

This Page Is Inserted by IFW Operations  
and is not a part of the Official Record

## **BEST AVAILABLE IMAGES**

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

**IMAGES ARE BEST AVAILABLE COPY.**

As rescanning documents *will not* correct images,  
please do not report the images to the  
Image Problem Mailbox.

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 09:12:13 ; Search time 33.7584 Seconds  
(without alignments)  
9093.534 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_5000\_6000  
Perfect score: 1001  
Sequence: 1 tgatctccctccccccacc.....tgggtcaccctctggtg 1001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents.NA.\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	52	5.2	7218	1	US-08-232-463-14
C 2	44.6	4.5	549	1	US-07-728-221B-12
C 3	44.6	4.5	549	3	US-07-728-220C-2
C 4	39.6	4.0	1894	3	US-07-728-220C-1
C 5	39	3.9	7218	1	US-08-232-463-14
C 6	38.4	3.8	2378	4	US-09-441-340-27
C 7	38	3.8	4032	4	US-09-068-101-5
C 8	37.4	3.7	5326	3	US-08-658-136-2
C 9	37.4	3.7	53577	3	US-08-658-136-1
C 10	36.6	3.7	2480	4	US-09-534-638-3
C 11	36.4	3.6	3070	1	US-08-428-732-3
C 12	36.4	3.6	111282	4	US-09-754-250-3
C 13	35.4	3.5	1969	4	US-09-098-628-1
C 14	34.8	3.5	320	4	US-09-165-264-7
C 15	34	3.4	320	4	US-09-165-264-13
C 16	34	3.4	1403	2	US-08-202-044-1
C 17	34	3.4	1403	4	US-08-751-344B-1
C 18	34	3.4	1614	1	US-08-583-672-1
C 19	33.8	3.4	75	4	US-09-461-697-151
C 20	33.8	3.4	320	4	US-09-165-264-14
C 21	33.8	3.4	517	1	US-08-480-784-33
C 22	33.8	3.4	517	1	US-08-483-553-33
C 23	33.8	3.4	517	1	US-08-487-002-33
C 24	33.8	3.4	517	1	US-08-483-554B-33
C 25	33.8	3.4	517	1	US-08-488-011B-33
C 26	33.8	3.4	517	4	US-08-850-727-33
C 27	33.8	3.4	517	5	PCT-US95-10202-33

Sequence 33, Appl  
Sequence 33, Appl  
Sequence 136, Appl  
Sequence 57, Appl  
Sequence 1, Appl  
Sequence 3, Appl  
Sequence 3, Appl  
Sequence 7, Appl  
Sequence 1, Appl  
Sequence 323, Appl  
Sequence 1, Appl  
Sequence 18, Appl  
Sequence 1, Appl  
Sequence 129, Appl  
Sequence 3, Appl  
Sequence 318, Appl

C 28 33.8 3.4 517 5 PCT-US95-10203-33  
C 29 33.8 3.4 517 5 PCT-US95-10220-33  
C 30 33.8 3.4 542 4 US-09-461-697-136  
C 31 33.8 3.4 1829 2 US-08-687-080-57  
C 32 33.8 3.4 8438 1 US-07-945-283-1  
C 33 33.6 3.4 8285 4 US-09-732-025-3  
C 34 33.6 3.4 11827 4 US-09-739-455-3  
C 35 33.6 3.4 44377 2 US-08-804-227C-7  
C 36 33.6 3.4 44377 2 US-08-804-198-1  
C 37 33.4 3.3 1166 4 US-09-072-596-323  
C 38 33.4 3.3 1422 1 US-08-439-725A-1  
C 39 33.4 3.3 1422 2 US-08-867-471-1  
C 40 33.4 3.3 1422 3 US-08-705-245-18  
C 41 33.4 3.3 1422 5 PCT-US96-0666A-1  
C 42 33.4 3.3 5590 4 US-09-050-159-129  
C 43 33.4 3.3 17138 4 US-09-813-819-3  
C 44 33.4 3.3 17138 4 US-09-920-048-3  
C 45 33.2 3.3 712 4 US-09-149-476-318

## ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14

Query Match

5.2%; Score 52; DB 1; Length 7218;



Best Local Similarity 4.3%; Pred. No. 2.7e-05;  
Matches 16; Conservative 206; Mismatches 146; Indels 0; Gaps 0;

QY 193 GGGAGAGATAGTGGGAGAGAAACAGCCAGCGGCTGGTGGTGGAGTGGAGAGCAGA 252  
Db 1406 RRR 1347  
QY 253 CGGCGTGGGAGGTCAGGAGAGATCTGCTGGGCTGGGATGGTGGGCAATCAACTGT 312  
Db 1346 RRR 1287  
QY 313 CCCATGCTGAGGCTGCTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 372  
Db 1286 RRR 1227  
QY 373 TCAGCCAGGCTGGCTGGGAGTGGTGGCCAGGCACTACTAAGAGCCAGGAAGCCCTGC 432  
Db 1226 RRR 1167  
QY 433 CAAGTGTGCTGGCTAGTTCCTCTGTCATCAGCCGCTAGCAGCCGCCCACTGTGCTGCAG 492  
Db 1166 RRR 1107  
QY 493 GTAAGGGGGAGGTTGAGCACATAGTACGCCCTGCTGTTCCATGCTTCCCTCT 552  
Db 1106 RRR 1047  
QY 553 GTGCCCCA 560  
Db 1046 GCAGCCAA 1039

RESULT 2  
US-07-728-221B-12/C  
; Sequence 12, Application US/07728221B  
; Patent No. 5340920  
; GENERAL INFORMATION:  
; APPLICANT: Matsuo, Hisayuki  
; APPLICANT: Kangawa, Kenji  
; APPLICANT: Tanaka, Shoji  
; APPLICANT: Fuchimura, Kayoko  
; APPLICANT: Tawaragi, Yasunori  
; TITLE OF INVENTION: NOVEL HYDROLOGICALLY ACTIVE PORCINE  
; NUMBER OF SEQUENCES: 12  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: ELEVENTH FLOOR, 1615 L STREET, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-5601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/728,221B  
; FILING DATE: 19910712  
; CLASSIFICATION: 530  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SCOTT, WATSON T.  
; REGISTRATION NUMBER: 26,581  
; REFERENCE/DOCKET NUMBER: WTS/9437/91817/KIK  
; TELEPHONE: (202) 861-3000  
; TELEFAX: (202) 822-0944  
; TELEX: 671 4627 CUSH  
; INFORMATION FOR SEQ ID NO: 12:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 549 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)  
US-07-728-221B-12

Query Match 4.5%; Score 44.6; DB 1; Length 549;  
Best Local Similarity 49.5%; Pred. No. 0.0013;  
Matches 145; Conservative 0; Mismatches 144; Indels 4; Gaps 1;

QY 89 TGGCTCACCCCTCTGCTCCCTCACCAGGCGCCCAACCGCCGCCAGCAG 148  
Db 306 TGTGCGCCCTCTCTGACCGCCCGCCAGCTGGGCTGGCCACCTCTCCCTCGCG 247  
QY 149 CTGTTCTCAGGCTCTCAGC---CTGCTGATTTCTGTTGCTGGGAGGAGATGAG 204  
Db 246 GAGTTCGAGGAGACCTTCGCGCGGCTTCGGGCTTGGCTTCGGAGGCGCGAGGAGGA 187  
QY 205 GTGGGAGAGAAACAGCCAGGCGAGGCTGTTGTTGAGTGAAGAGCAGACGCGGTTGGGA 264  
Db 186 GCGTGAAGAGAGAGAGCGGAGGCGGAGCTGGGAGGAGGTGCATGGTGGCTGTGGGTTGA 127  
QY 265 GGTGAGGAGAGATCTGCTGGGCTGGGATGGTGGGATCAACTGTCCCATTTGCTGCA 324  
Db 126 GGGGCGAGATCGGCGGAGAGAGGTGCACGGGCGCGGCGGCGAGACTGCCAAGCGGCG 67  
QY 325 GCGTGTCTTGGGCGAGGAGGAGGATGGGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGG 377  
Db 66 CGGAGCAGACTGCGCGGCTGCGGTGCGGCGGCGGAGTCCAGTGTCTGCGCGGC 14

RESULT 3  
US-07-728-220C-2/C  
; Sequence 2, Application US/07728220C  
; Patent No. 6020168  
; GENERAL INFORMATION:  
; APPLICANT: MATSUO, HISAYUKI  
; APPLICANT: KANGAWA, KENJI  
; APPLICANT: TANAKA, SHOJI  
; APPLICANT: FUCHIMURA, KAYOKO  
; APPLICANT: TAWARAGI, YASUNORI  
; TITLE OF INVENTION: PORCINE CNP GENE AND PRECURSOR PROTEIN  
; NUMBER OF SEQUENCES: 22  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
; STREET: ELEVENTH FLOOR, 1615 L STREET, N.W.  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20036-5601  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/728,220C  
; FILING DATE: 19910712  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: SCOTT, WATSON T.  
; REGISTRATION NUMBER: 26,581  
; REFERENCE/DOCKET NUMBER: WTS/9437/91816  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)861-3000  
; TELEFAX: (202)822-0944  
; TELEX: 671 4627 CUSH  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 549 base pairs  
; TYPE: NUCLEIC ACID  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
US-07-728-220C-2









```

1  CITY: Palo Alto
2  STATE: CA
3  COUNTRY: USA
4  ZIP: 94304-1018
5  COMPUTER READABLE FORM:
6  MEDIUM TYPE: Diskette
7  COMPUTER: IBM Compatible
8  OPERATING SYSTEM: Windows
9  SOFTWARE: FastSeq for Windows Version 2.0b
10 CURRENT APPLICATION DATA:
11 APPLICATION NUMBER: US/09/098,628
12 FILING DATE:
13 CLASSIFICATION:
14 PRIOR APPLICATION DATA:
15 APPLICATION NUMBER:
16 FILING DATE:
17 ATTORNEY/AGENT INFORMATION:
18 NAME: Lehnhardt, Susan K
19 REGISTRATION NUMBER: 33,943
20 REFERENCE/DOCKET NUMBER: 21900-20089.10
21 TELECOMMUNICATION INFORMATION:
22 TELEPHONE: 650-813-5600
23 TELEFAX: 650-494-0792
24 TELEX: 706141
25 INFORMATION FOR SEQ ID NO: 1:
26 SEQUENCE CHARACTERISTICS:
27 LENGTH: 1969 base pairs
28 TYPE: nucleic acid
29 STRANDEDNESS: double
30 TOPOLOGY: linear
31 MOLECULE TYPE: cDNA
32 FEATURE:
33 NAME/KEY: Coding Sequence
34 LOCATION: 970...1434
35 OTHER INFORMATION:
36 US-09-098-628-1

```

	Query Match	3.5%;	Score 35.4;	DB 4;	Length 1969;
	Best Local Similarity	50.3%;	Pred. No. 1;	Mismatches	87; Conservative
					0; Gaps
QY	94	CACCCCTCTGCTTCCTGGCCTCA	CCCCAGGCCCGCCCCAACCCGCGCCCCCCCAGCAGCTGT	TT	153
Db	376	CCCTGCCGCGGCCTCCGACGGCGCGCCCCCGCCCTCTGCCCGCCCCCGACTGA			435
QY	154	CTCAGGCCCTCTACGCCCTGTCTGATTGCTTCTCTGGCCTGGGAGAAATGAGTG	GAGAA		213
Db	436	GGCGGCGCTCCCCCGCGACTGATGTCGCGCGCTTCGCTGTGTGTGGCGAAGCCCGCAA			495
QY	214	AACCAGCCAGGGCAGTTGGTTGGAGTGAAGACAGACCGCGGTGGGGAGG			266
Db	496	CTCAGAGCCCGCGCCCCAGAAAACCCAGCGAGTAGGGGGCGCGCGCAGGAGG			548

```

RESULT 14
US-09-165-264-7
; Sequence 7, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vineyagamoorthy, Thuralaayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 7
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; - OTHER INFORMATION: Description of Artificial Seq
US-09-165-264-7

```

[illegible][illegible]

Search completed: March 30, 2003, 13:56:22  
Job time : 174.758 secs

PR	14-AUG-2000;	2000US-02254447.
PR	14-AUG-2000;	2000US-02255757.
PR	14-AUG-2000;	2000US-02255758.
PR	18-AUG-2000;	2000US-02257759.
PR	22-AUG-2000;	2000US-022626279.
PR	22-AUG-2000;	2000US-0226681.
PR	22-AUG-2000;	2000US-0226868.
PR	23-AUG-2000;	2000US-0227182.
PR	30-AUG-2000;	2000US-0227009.
PR	01-SEP-2000;	2000US-0228924.
PR	01-SEP-2000;	2000US-0229287.
PR	01-SEP-2000;	2000US-0229343.
PR	01-SEP-2000;	2000US-0229344.
PR	05-SEP-2000;	2000US-0229345.
PR	05-SEP-2000;	2000US-0229509.
PR	06-SEP-2000;	2000US-0229513.
PR	06-SEP-2000;	2000US-0230437.
PR	08-SEP-2000;	2000US-0230438.
PR	08-SEP-2000;	2000US-0231242.
PR	08-SEP-2000;	2000US-0231243.
PR	08-SEP-2000;	2000US-0231244.
PR	08-SEP-2000;	2000US-0231413.
PR	08-SEP-2000;	2000US-0231414.
PR	08-SEP-2000;	2000US-0232080.
PR	12-SEP-2000;	2000US-0232081.
PR	14-SEP-2000;	2000US-0231968.
PR	14-SEP-2000;	2000US-0232397.
PR	14-SEP-2000;	2000US-0232398.
PR	14-SEP-2000;	2000US-0232399.
PR	14-SEP-2000;	2000US-0232400.
PR	14-SEP-2000;	2000US-0232401.
PR	14-SEP-2000;	2000US-0232401.
PR	14-SEP-2000;	2000US-0233063.
PR	14-SEP-2000;	2000US-0233064.
PR	21-SEP-2000;	2000US-0233065.
PR	21-SEP-2000;	2000US-0233223.
PR	25-SEP-2000;	2000US-0234274.
PR	25-SEP-2000;	2000US-0234997.
PR	26-SEP-2000;	2000US-0234998.
PR	27-SEP-2000;	2000US-0235484.
PR	27-SEP-2000;	2000US-0235834.
PR	29-SEP-2000;	2000US-0235836.
PR	29-SEP-2000;	2000US-0236327.
PR	29-SEP-2000;	2000US-0236367.
PR	29-SEP-2000;	2000US-0236368.
PR	29-SEP-2000;	2000US-0236369.
PR	02-OCT-2000;	2000US-0236370.
PR	02-OCT-2000;	2000US-0236802.
PR	02-OCT-2000;	2000US-0237037.
PR	02-OCT-2000;	2000US-0237038.
PR	02-OCT-2000;	2000US-0237039.
PR	13-OCT-2000;	2000US-0237040.
PR	13-OCT-2000;	2000US-0239935.
PR	20-OCT-2000;	2000US-0239937.
PR	20-OCT-2000;	2000US-0240960.
PR	20-OCT-2000;	2000US-0241221.
PR	20-OCT-2000;	2000US-0241785.
PR	20-OCT-2000;	2000US-0241786.
PR	20-OCT-2000;	2000US-0241787.
PR	20-OCT-2000;	2000US-0241808.
PR	20-OCT-2000;	2000US-0241809.
PR	01-NOV-2000;	2000US-0241826.
PR	08-NOV-2000;	2000US-0244617.
PR	08-NOV-2000;	2000US-0246474.
PR	08-NOV-2000;	2000US-0246475.
PR	08-NOV-2000;	2000US-0246476.
PR	08-NOV-2000;	2000US-0246477.
PR	08-NOV-2000;	2000US-0246478.
PR	08-NOV-2000;	2000US-0246523.
PR	08-NOV-2000;	2000US-0246524.
PR	08-NOV-2000;	2000US-0246525.
PR	08-NOV-2000;	2000US-0246526.
PR	08-NOV-2000;	2000US-0246527.
PR	08-NOV-2000;	2000US-0246528.

PR	08-NOV-2000;	2000US-0245532.
PR	08-NOV-2000;	2000US-0246609.
PR	08-NOV-2000;	2000US-0246610.
PR	08-NOV-2000;	2000US-0246611.
PR	08-NOV-2000;	2000US-0246613.
PR	17-NOV-2000;	2000US-0249207.
PR	17-NOV-2000;	2000US-0249208.
PR	17-NOV-2000;	2000US-0249209.
PR	17-NOV-2000;	2000US-0249210.
PR	17-NOV-2000;	2000US-0249211.
PR	17-NOV-2000;	2000US-0249212.
PR	17-NOV-2000;	2000US-0249213.
PR	17-NOV-2000;	2000US-0249214.
PR	17-NOV-2000;	2000US-0249215.
PR	17-NOV-2000;	2000US-0249216.
PR	17-NOV-2000;	2000US-0249217.
PR	17-NOV-2000;	2000US-0249218.
PR	17-NOV-2000;	2000US-0249244.
PR	17-NOV-2000;	2000US-0249245.
PR	17-NOV-2000;	2000US-0249264.
PR	17-NOV-2000;	2000US-0249265.
PR	17-NOV-2000;	2000US-0249267.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	05-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-02559678.
XX		
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX		
PI	Rosen CA, Barash SC, Ruben SM;	
DR	WPI; 2001-488786/53.	
XX		
PT	New isolated ovarian and/or breast cancer related nucleic acids and	
PT	polypeptides, useful for diagnosing, treating and/or preventing human	
PT	diseases and disorders, particularly ovarian and/or breast cancer -	
XX		
XX	Disclosure; SEQ ID NO 1003; 577pp + Sequence Listing; English.	
PS		
XX		
CC	The invention relates to novel genes (ABA07454-ABA08224) and proteins	
CC	(AB010743-AB010980) useful for preventing, treating or ameliorating	
CC	medical conditions e.g. by protein or gene therapy. The genes are	
CC	isolated from a range of human tissues disclosed in the specification.	
CC	The nucleic acids, proteins, antibodies and (antagonists are useful	
CC	in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast	
CC	and ovarian cancer and other cancers of the adrenal gland, bone, bone	
CC	marrow, breast, gastrointestinal tract, liver, lung, or urogenital;	
CC	(b) immune disorders e.g. Addison's disease, allergies, autoimmune	
CC	haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's	
CC	disease, multiple sclerosis, rheumatoid arthritis and ulcerative	
CC	colitis; (c) cardiovascular disorders such as myocardial ischaemias;	
CC	(d) wound healing; (e) neurological diseases e.g. cerebral anoxia and	
CC	epilepsy; and (f) infectious diseases such as viral, bacterial, fungal	
CC	and parasitic infections.	
CC	Note: the sequence data for this patent did not form part of the	
CC	printed specification, but was obtained in electronic format directly	
CC	from WIPO at ftp.wipo.int/pub/published_int_sequences.	
XX		
XX	Sequence 13862 BP; 3188 A; 3560 C; 4465 G; 2649 T; 0 other;	
SQ		
Query Match	5.4%;	Score 54; DB 22; Length 13862;
Best Local Similarity	47.0%;	Pred. No. 0.013;

	Matches	232;	Conservative	0;	Mismatches	260;	Indels	2;	Gaps	2;
QY	461	CGCGGACCCGGGCCCTAGCCTCTCTGCGCGCGCTCTCGCGCTCTCCGGGGGAGGTGGGCC	520							
DB	6784	CCCGTCAACGCCTTCAGCGCCCCGACACTCCACCACCCCGGAGATGGTT	5843							
QY	521	GGTTCGGCCGGGACGGGGCTTGCGGGGCGAGCCCOCGGGGGGGTGCGGAGCGGGTGAT	580							
DB	6844	TCCAGCAACACAGCGGCTTTGGTGAGGACGGGCCCTATGGGGCGGAGGGGCGGGAG	6903							
QY	581	GTCAGGGCACGGGTGGGTGGGTCACTCGGAGGTGAGGCGCGCCGACGGGAGTTCAGCGA	640							
DB	6904	CCATGGGGCGACGGGGCGGGGTCTATGGGTGGGAGGGGCTTGGTCGGGAGGGCGGGA	6963							
QY	641	GAGTTCAGCCCGCATTTAGGCAAAATGAGCGCCCGGCTGGGTGGGGTGTGTAAAG	700							
DB	6964	CCAGAGGSCGGAGTGGCGAGGTCTGACAGCTGACCGCGGTCTTTCCTTCTGCCGCA	7023							
QY	701	GGGAGGACAC-CCGGACACCCCCCTCTTCCCGCGCCACCACTCTCCACACAGGGTT	759							
DB	7024	GGAGCAGTTTCGAGTGGGCGGAGGATTTCGCCCTCTCCGCCCTCCAGGCCGCCCGCTG	7083							
QY	760	CGCTCGGCCAGGAGCTGACCAAACCTTGGGGGAGCGCTGGGAGCGGAACGTACAAGG	819							
DB	7084	TGCTTCTCCCGCTTCTCGGTCTCGCTGAGACCCGGGGCCACCGCCCGGCC	7143							
QY	820	GAGGACGCCCGCCCTCTTTCGGTCTCTGTCCTCCCTCGAGCGCCCTCTCTCTCCCTGTACTC	879							
DB	7144	CCGAGCGCCGCGCGCAGGTACCATATCTCTCGGCCCTCGCCCTCTCTGCCCCATCTCT	7203							
QY	880	GGGTCCCTCTGTACTCTGTGTACTCTCATCTCTGGAGCCTTTCGCCCTTCTCTCTCTCT	939							
DB	7204	GTCCCCCTATTTCCACTGCCAACCTTCTGTGACCA-TCITTTCAATGTCCCTGCCTCTCT	7262							
QY	940	CCTCTCTCTCCCTCT 953								
DB	7263	CTTCCAGTCTCT 7276								
 RESULT 11 AAL02789 ID AAL02789 standard; DNA; 13862 BP. XX AC AAL02789; XX DT XX DT XX DE Human reproductive system related antigen DNA SEQ ID NO: 5477. XX KW Human; reproductive system .related antigen; reproductive system disorder; XX KW cancer; gene therapy; ds. XX OS Homo sapiens. XX PN W020015320-A2. XX PD 02-AUG-2001. XX PF 17-JAN-2001; 2001WO-US01339. XX PR 31-JAN-2000; 2000US-0179065. PR 04-FEB-2000; 2000US-0180628. PR 24-FEB-2000; 2000US-0184664. PR 02-MAR-2000; 2000US-0186350. PR 16-MAR-2000; 2000US-0189874. PR 17-MAR-2000; 2000US-0190076. PR 18-APR-2000; 2000US-0198123. PR 19-MAY-2000; 2000US-0205515. PR 07-JUN-2000; 2000US-0209467. PR 28-JUN-2000; 2000US-0214886. PR 30-JUN-2000; 2000US-0215135. PR 07-JUL-2000; 2000US-0216647. PR 07-JUL-2000; 2000US-0216880.										



PR 11-JUL-2000; 2000US-0217487.  
 PR 11-JUL-2000; 2000US-0217496.  
 PR 14-JUL-2000; 2000US-0218290.  
 PR 26-JUL-2000; 2000US-0220963.  
 PR 26-JUL-2000; 2000US-0220964.  
 PR 14-AUG-2000; 2000US-0224518.  
 PR 14-AUG-2000; 2000US-0224519.  
 PR 14-AUG-2000; 2000US-0225213.  
 PR 14-AUG-2000; 2000US-0225214.  
 PR 14-AUG-2000; 2000US-0225216.  
 PR 14-AUG-2000; 2000US-0225266.  
 PR 14-AUG-2000; 2000US-0225287.  
 PR 14-AUG-2000; 2000US-0225268.  
 PR 14-AUG-2000; 2000US-0225270.  
 PR 14-AUG-2000; 2000US-0225447.  
 PR 14-AUG-2000; 2000US-0225757.  
 PR 14-AUG-2000; 2000US-0225758.  
 PR 14-AUG-2000; 2000US-0225759.  
 PR 18-AUG-2000; 2000US-0226279.  
 PR 22-AUG-2000; 2000US-0226681.  
 PR 22-AUG-2000; 2000US-0226686.  
 PR 22-AUG-2000; 2000US-0227182.  
 PR 23-AUG-2000; 2000US-0227009.  
 PR 30-AUG-2000; 2000US-0228924.  
 PR 01-SEP-2000; 2000US-0229287.  
 PR 01-SEP-2000; 2000US-0229343.  
 PR 01-SEP-2000; 2000US-0229344.  
 PR 01-SEP-2000; 2000US-0229345.  
 PR 05-SEP-2000; 2000US-0229509.  
 PR 05-SEP-2000; 2000US-0229513.  
 PR 06-SEP-2000; 2000US-0230437.  
 PR 06-SEP-2000; 2000US-0230438.  
 PR 08-SEP-2000; 2000US-0231242.  
 PR 08-SEP-2000; 2000US-0231243.  
 PR 08-SEP-2000; 2000US-0231244.  
 PR 08-SEP-2000; 2000US-0231413.  
 PR 08-SEP-2000; 2000US-0231414.  
 PR 08-SEP-2000; 2000US-0232080.  
 PR 08-SEP-2000; 2000US-0232081.  
 PR 12-SEP-2000; 2000US-0231968.  
 PR 14-SEP-2000; 2000US-0232397.  
 PR 14-SEP-2000; 2000US-0232398.  
 PR 14-SEP-2000; 2000US-0232399.  
 PR 14-SEP-2000; 2000US-0232400.  
 PR 14-SEP-2000; 2000US-0232401.  
 PR 14-SEP-2000; 2000US-0233063.  
 PR 14-SEP-2000; 2000US-0233064.  
 PR 14-SEP-2000; 2000US-0233065.  
 PR 21-SEP-2000; 2000US-0234223.  
 PR 21-SEP-2000; 2000US-0234274.  
 PR 25-SEP-2000; 2000US-0234997.  
 PR 25-SEP-2000; 2000US-0234998.  
 PR 26-SEP-2000; 2000US-0235484.  
 PR 27-SEP-2000; 2000US-0235834.  
 PR 27-SEP-2000; 2000US-0235836.  
 PR 29-SEP-2000; 2000US-0236327.  
 PR 29-SEP-2000; 2000US-0236367.  
 PR 29-SEP-2000; 2000US-0236368.  
 PR 29-SEP-2000; 2000US-0236369.  
 PR 29-SEP-2000; 2000US-0236370.  
 PR 02-OCT-2000; 2000US-0236802.  
 PR 02-OCT-2000; 2000US-0237037.  
 PR 02-OCT-2000; 2000US-0237038.  
 PR 02-OCT-2000; 2000US-0237039.  
 PR 02-OCT-2000; 2000US-0237040.  
 PR 13-OCT-2000; 2000US-0239935.  
 PR 13-OCT-2000; 2000US-0239937.  
 PR 20-OCT-2000; 2000US-0240960.  
 PR 20-OCT-2000; 2000US-0241221.  
 PR 20-OCT-2000; 2000US-0241785.  
 PR 20-OCT-2000; 2000US-0241786.  
 PR 20-OCT-2000; 2000US-0241787.  
 PR 20-OCT-2000; 2000US-0241808.  
 PR 20-OCT-2000; 2000US-0241809.

PR 20-OCT-2000; 2000US-0241826.  
 PR 01-NOV-2000; 2000US-0244617.  
 PR 08-NOV-2000; 2000US-0244674.  
 PR 08-NOV-2000; 2000US-0244675.  
 PR 08-NOV-2000; 2000US-0244676.  
 PR 08-NOV-2000; 2000US-0244677.  
 PR 08-NOV-2000; 2000US-0244678.  
 PR 08-NOV-2000; 2000US-024523.  
 PR 08-NOV-2000; 2000US-024524.  
 PR 08-NOV-2000; 2000US-024525.  
 PR 08-NOV-2000; 2000US-024526.  
 PR 08-NOV-2000; 2000US-024527.  
 PR 08-NOV-2000; 2000US-024528.  
 PR 08-NOV-2000; 2000US-024532.  
 PR 08-NOV-2000; 2000US-024532.  
 PR 08-NOV-2000; 2000US-0245609.  
 PR 08-NOV-2000; 2000US-0246610.  
 PR 08-NOV-2000; 2000US-0246611.  
 PR 08-NOV-2000; 2000US-0246613.  
 PR 17-NOV-2000; 2000US-0249207.  
 PR 17-NOV-2000; 2000US-0249208.  
 PR 17-NOV-2000; 2000US-0249209.  
 PR 17-NOV-2000; 2000US-0249210.  
 PR 17-NOV-2000; 2000US-0249211.  
 PR 17-NOV-2000; 2000US-0249212.  
 PR 17-NOV-2000; 2000US-0249213.  
 PR 17-NOV-2000; 2000US-0249214.  
 PR 17-NOV-2000; 2000US-0249215.  
 PR 17-NOV-2000; 2000US-0249216.  
 PR 17-NOV-2000; 2000US-0249217.  
 PR 17-NOV-2000; 2000US-0249218.  
 PR 17-NOV-2000; 2000US-0249244.  
 PR 17-NOV-2000; 2000US-0249245.  
 PR 17-NOV-2000; 2000US-0249264.  
 PR 17-NOV-2000; 2000US-0249265.  
 PR 17-NOV-2000; 2000US-0249297.  
 PR 17-NOV-2000; 2000US-0249299.  
 PR 01-DEC-2000; 2000US-0249300.  
 PR 01-DEC-2000; 2000US-0250160.  
 PR 01-DEC-2000; 2000US-0250391.  
 PR 05-DEC-2000; 2000US-0251030.  
 PR 05-DEC-2000; 2000US-0251988.  
 PR 05-DEC-2000; 2000US-0256719.  
 PR 06-DEC-2000; 2000US-0251479.  
 PR 08-DEC-2000; 2000US-0251856.  
 PR 08-DEC-2000; 2000US-0251868.  
 PR 08-DEC-2000; 2000US-0251869.  
 PR 08-DEC-2000; 2000US-0251989.  
 PR 08-DEC-2000; 2000US-0251990.  
 PR 11-DEC-2000; 2000US-0254097.  
 PR 05-JAN-2001; 2001US-0259678.  
 XX (HUMA-) HUMAN GENOME SCI INC.  
 PA Rosen CA, Barash SC, Ruben SM;  
 PI WPI; 2001-465570/50.  
 DR Isolated nucleic acid molecule encoding a reproductive system antigen  
 XX is used in preventing, treating or ameliorating a medical condition -  
 PT Disclosure; SEQ ID NO 5477; 1297pp + Sequence Listing; English.  
 PS  
 XX  
 CC The present invention provides the protein and coding sequences of a  
 CC number of human reproductive system related antigens. These can be used  
 CC in the prevention and treatment of reproductive system disorders,  
 CC including cancer. The present sequence is a genomic sequence encoding a  
 CC protein of the invention.  
 XX  
 SQ Sequence 13862 BP; 3188 A; 3560 C; 4465 G; 2649 T; 0 other;  
 Query Match 5.4%; Score 54; DB 22; Length 13862;  
 Best Local Similarity 47.0%; Pred. No. 0.013;  
 Matches 232; Conservative 0; Mismatches 260; Indels 2; Gaps 2;









GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: March 30, 2003, 09:12:13 ; Search time 33.7247 Seconds  
(without alignments)  
9093.534 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_1\_1000  
Perfect score: 1000  
Sequence: 1 gtcctgggtagcatgtaca.....acatgcgcctctccctctc 1000

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents\_NA: \*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq: \*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq: \*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq: \*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq: \*  
5: /cgn2\_6/ptodata/1/ina/PTUS\_COMB.seq: \*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	72.6	7.3	7218	1	US-08-232-463-14		Sequence 14, Appl
2	59.2	5.9	53526	3	US-08-658-136-2		Sequence 2, Appl
3	59.2	5.9	53577	3	US-08-658-136-1		Sequence 1, Appl
4	51.6	5.2	4897	6	5196516-7		Patent No. 5196516
C 5	49.6	5.0	320	4	US-09-165-264-7		Sequence 7, Appl
C 6	49.6	5.0	320	4	US-09-165-264-14		Sequence 14, Appl
C 7	49.2	4.9	318	4	US-09-165-264-12		Sequence 12, Appl
C 8	48.6	4.9	319	4	US-09-165-264-8		Sequence 8, Appl
C 9	48.6	4.9	3978	3	US-08-726-214-1		Sequence 1, Appl
C 10	48.4	4.8	152331	3	US-09-128-155-16		Sequence 16, Appl
C 11	48	4.8	320	4	US-09-165-264-11		Sequence 11, Appl
C 12	47.8	4.8	320	4	US-09-165-264-13		Sequence 13, Appl
13	44	4.4	320	4	US-09-165-264-13		Sequence 13, Appl
14	43.8	4.4	319	4	US-09-165-264-8		Sequence 8, Appl
C 15	43.2	4.3	32207	2	US-08-770-379-20		Sequence 20, Appl
C 16	43.2	4.3	32207	4	US-08-757-669A-20		Sequence 20, Appl
C 17	43.2	4.3	32207	4	US-09-230-371A-20		Sequence 20, Appl
18	42.8	4.3	152331	3	US-09-128-155-16		Sequence 16, Appl
C 19	42.6	4.3	3978	3	US-08-726-214-1		Sequence 1, Appl
20	42	4.2	80248	4	US-09-078-294-4		Sequence 4, Appl
21	42	4.2	80595	4	US-09-078-294-3		Sequence 3, Appl
22	41.8	4.2	320	4	US-09-165-264-7		Sequence 7, Appl
23	41.4	4.1	320	4	US-09-165-264-11		Sequence 11, Appl
24	41.4	4.1	320	4	US-09-165-264-14		Sequence 14, Appl
25	41.4	4.1	950	4	US-08-931-789A-6		Sequence 6, Appl
26	41.4	4.1	950	4	US-09-062-451-6		Sequence 6, Appl
27	41.4	4.1	950	4	US-09-598-326-6		Sequence 6, Appl

28	41.2	4.1	1192	5	PCT-US93-06251-51	Sequence 51, Appl
29	41.2	4.1	1320	3	US-08-989-251-28	Sequence 28, Appl
30	41.2	4.1	1320	3	US-09-340-250-28	Sequence 28, Appl
31	41.2	4.1	1320	4	US-09-528-108-28	Sequence 28, Appl
C 32	41.2	4.1	4496	4	US-08-765-907A-6	Sequence 6, Appl
33	41	4.1	1500	4	US-09-593-711A-10	Sequence 10, Appl
C 34	40.8	4.1	289	4	US-09-007-005-17	Sequence 17, Appl
C 35	40.8	4.1	289	4	US-09-244-796-17	Sequence 17, Appl
C 36	40.4	4.0	801	2	US-08-770-379-16	Sequence 16, Appl
C 37	40.4	4.0	801	4	US-08-757-669A-16	Sequence 16, Appl
C 38	40.4	4.0	801	4	US-09-230-371A-16	Sequence 16, Appl
C 39	40.2	4.0	318	4	US-09-165-264-12	Sequence 12, Appl
40	40.2	4.0	1877	4	US-09-780-173A-10	Sequence 10, Appl
41	40	4.0	135	4	US-09-018-635-36	Sequence 36, Appl
42	40	4.0	801	2	US-08-770-379-16	Sequence 16, Appl
43	40	4.0	801	4	US-08-757-669A-16	Sequence 16, Appl
44	40	4.0	801	4	US-09-230-371A-16	Sequence 16, Appl
45	40	4.0	2907	2	US-09-018-628-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1  
US-08-232-463-14  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHEIFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703)836-9300  
; TELEFAX: (703)683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: ptzqpt-Fls  
; US-08-232-463-14

Query Match 7.3%; Score 72.6; DB 1; Length 7218;



```
RESULT 4
5196516-7
; APPLICANT: SCHREURS, CHRISTA S.;METTENLEITER, THOMAS C.
; SIMON, ARTUR J.;LURAS, NOEMI;RZIHA, HANNS J.
; TITLE OF INVENTION: PSEUDORABIES VIRUS VACCINE
; NUMBER OF SEQUENCES: 8
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/383.833
; FILING DATE: 21-JUL-1989
; SEQ ID NO:7:
; LENGTH: 4897
5196516-7

Query Match          5.2%; Score 51.6; DB 6; Length 4897;
Best Local Similarity 55.6%; Pred. No. 0.001;
Matches 99; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 823 GACGCCGCCCTCTCGCTCTTGTCCCTCGCAGCCCTCTCTCCCTGCTACTCGGC 882
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3900.GTCCCTCTCCCTCGCTCTCTCCCTCGCTCTCTCCCTCGCTCTCTCCCTCTCC 3959

QY 883 GTCCCTCTGTACTCTGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 942
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3960 GTCCCTCTCTCCCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4019

QY 943 CTCCTCCCTCTCCAGCTGCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1000
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 4020 CTCGCTCTCTCTCCCTCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 4077

RESULT 5
US-09-165-264-7/c
; Sequence 7, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 7
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-7

Query Match          5.0%; Score 49.6; DB 4; Length 320;
Best Local Similarity 47.7%; Pred. No. 0.0013;
Matches 145; Conservative 0; Mismatches 159; Indels 0; Gaps 0;

QY 697 TAAGGGGAGGACACCGGGACACCCCTCTTCCCGCCGCCACCACTCTCTCCACG 756
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 318 TAAGGGCATGGCTAGCTCTCCCGCCGCCACCCCGCCGCCCGCCCGCCCGCC 259

QY 757 CTTGCTGGCCAGGGACTGACCAAACTTTGGGGAGCCTGGGAGCGGAAGTGTACAA 816
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 258 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 199

QY 817 GGGGAGGACCGCCCGCTCTTGGTCTCTTGTCCCTCGAGCCCGCTCTCTCTCT 876
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 198 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 139

QY 877 CTGGGCTCTCTGTACTCTGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 936
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 138 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 79

QY 937 TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 996
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
DB 78 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 19
QY 997 TCTC 1000
      |||
DB 18 CCCC 15

RESULT 6
US-09-165-264-14/c
; Sequence 14, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-14

Query Match          5.0%; Score 49.6; DB 4; Length 320;
Best Local Similarity 48.0%; Pred. No. 0.0013;
Matches 142; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 705 GGACACCGGGACACCCCTCTTCCCGCCGCCACCACTCTCTCCACCGCTTCTCT 764
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 305 GCACACCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 246

QY 765 GGCCAGGAGACTGACCAAACTTTGGGGAGCCTGGGAGCGGAAGTGTACAAAG 824
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 245 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 186

QY 825 CGCCCGCCCGCTTCTGGTCTCTTGTCCCTCGAGCCCGCTCTCTCTCTCTCT 884
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 185 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 126

QY 885 CCCTCTGTACTCTGTGTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 944
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 125 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 66

QY 945 COTCCCGCTTCCAGGCTGCCCCCACTTCCCTGTCCACATGCGCGCTCTCTCT 1000
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 65 CCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 10

RESULT 7
US-09-165-264-12/c
; Sequence 12, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuralayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 12
; LENGTH: 318
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-12

Query Match          4.9%; Score 49.2; DB 4; Length 318;
```



```

Best Local Similarity 48.0%; Pred. No. 0.0017;
Matches 141; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 707 ACACCGGACACCCCTCTTCCCGCGCCACACCTCTCTCCACACGGCTTCGCTCG 766
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 309 ACACACCAACCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 250
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 767 CCAGGACACTGACAAACCTTGGGGAGCCTGGGAGCCGGAATGTTACAAGGGAGGACG 826
    || || || || || || || || || || || || || || || || || || ||
Db 249 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 190
    || || || || || || || || || || || || || || || || || || ||
QY 827 CCGCGCCCTCTTCGCTTCCTGCTCGAGCCGCTCTCTCTCTCTCTCTCTCTCT 886
    || || || || || || || || || || || || || || || || || || ||
Db 189 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 130
    || || || || || || || || || || || || || || || || || || ||
QY 887 CTCTGTACTCTGTACTCTCATCTGGAGCCTTTCGCCCTTCCTGCTTCTCTCTCT 946
    || || || || || || || || || || || || || || || || || || ||
Db 129 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 70
    || || || || || || || || || || || || || || || || || || ||
QY 947 TCCCGCTTCCAGGTGCGCCGACCTGCTGTCACATGCGCGCTCTCTCTCTCT 1000
    || || || || || || || || || || || || || || || || || || ||
Db 69 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 16

RESULT 8
US-09-165-264-8/C
; Sequence 8, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiyavah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ. ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 319
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-8

Query Match 4.9%; Score 48.6; DB 4; Length 319;
Best Local Similarity 47.8%; Pred. No. 0.0024;
Matches 141; Conservative 0; Mismatches 154; Indels 0; Gaps 0;

QY 706 GACACGGGACACCCCGCTTTCGCGCCACACCTCTCTCCACACGGCTTCGCTCG 765
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 302 GACCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 243
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 766 GCACGGGACTGACAAACCTTGGGGAGCCTGGGAGCCGGAATGTTACAAGGGAGGAC 825
    || || || || || || || || || || || || || || || || || || ||
Db 242 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 183
    || || || || || || || || || || || || || || || || || || ||
QY 826 GCGCGCCCTCTTTCGCTTCCTGCTCGAGCCGCTCTCTCTCTCTCTCTCTCTCT 885
    || || || || || || || || || || || || || || || || || || ||
Db 182 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 123
    || || || || || || || || || || || || || || || || || || ||
QY 886 CTCTGTACTCTGTACTCTCATCTGGAGCCTTTCGCCCTTCCTGCTTCTCTCTCT 945
    || || || || || || || || || || || || || || || || || || ||
Db 122 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 63
    || || || || || || || || || || || || || || || || || || ||
QY 946 CTCCCGCTTCCAGGTGCGCCGACCTGCTGTCACATGCGCGCTCTCTCTCTCTCT 1000
    || || || || || || || || || || || || || || || || || || ||
Db 62 CCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG 8
    || || || || || || || || || || || || || || || || || || ||

```

```

; GENERAL INFORMATION:
; APPLICANT: Tang, Wei-Jen
; APPLICANT: Gilman, Alfred G.
; TITLE OF INVENTION: SOLUBLE MAMMALIAN ADENYLYL CYCLASE
; TITLE OF INVENTION: AND USES THEREOF
; NUMBER OF SEQUENCES: 31
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: Concurrently Herewith
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/005,498
; FILING DATE: 04-OCT-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Highlander, Steven L.
; REGISTRATION NUMBER: 37,642
; REFERENCE/DOCKET NUMBER: UTSD:450
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 418-3000
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3978 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
;
US-08-726-214-1

Query Match 4.9% Score 48.6; DB 3; Length 3978;
Best Local Similarity 58.0%; Pred. No. 0.0053;
Matches 105; Conservative 0; Mismatches 74; Indels 2; Gaps

Qy 452 ACTTGGCCCGCGGACCCGCGCCTACGCCTCTCTGCC- -GCCGCTCTCGCGCTCTCCGG 509
    ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 13 ACTCGCCCGCGCGCGCCCGCGCCTCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCG 72
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

Qy 510 GGAGTGGCCGGTTCGGCGCGGAGGGGCTGCGGGCGAGCCCGCGCGCGGGCTGGC 569
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 73 CGAGGGCGCGCGCGCTGCGTGGATGGCGGGGCGCGCGCGCGCGCGAGCGCGC 132
    |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||

Qy 570 GAGCGGTGATGTCACGGGACGGTGGTGGTGTACTCGAGGTGAGCGCGCGCCAGCG 629
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 133 GCGGAGGCGGAGCGCGCGCGCGCGAGTCTGGGGCGCGAGCGGGCGCGCGCGCGC 192
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Qy 630 G 630
    |
Db 193 G 193

RESULT 10
US-09-128-155-16/c
; Sequence 16, Application US/09128155
; Patent No. 6117654
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/09/128,155
; EARLIER FILING DATE: 1998-08-03
; EARLIER APPLICATION NUMBER: US 60/091,650

```

```

; EARLIER FILING DATE: 1998-07-02
; EARLIER APPLICATION NUMBER: US 60/054,646
; EARLIER FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(152331)
; OTHER INFORMATION: n = A,T,C or G
US-09-128-155-16

```

Query Match	4.8%;	Score 48.4;	DB 3;	Length 152331;
Best Local Similarity	51.7%;	Pred. No. 0.018;		
Matches 134;	Conservative 0;	Mismatches 122;	Indels 3;	Gaps 1;
<hr/>				
QY	456	GGCCCCCGGACCCGCCCTACGCCTCCTCCCGCGGTCTCCGGGTCTCCGGGGAGGT	515	
Db	22139	GGCGCGCGGGGGGGGCGCTCGCTGCGGGGGGGGGCGCGGGCGTTCGGGTGGG	22080	
QY	516	GGCCCGGTTCGGCGCGCACAGGGGGCTGGCGGGCAGCCCCCGGGCGGGGTTCGGCAGCGG	575	
Db	22079	GGTGCGTGTGGCGCGGTGTGGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGG	22020	
QY	576	GTGATGTACAGGGCAGCGGTGGGTGGTCACTCCGAGGTGAGGGCCGCCAGGGCGAGTTC	635	
Db	22019	GTGGGGG---GCGCGGGGGGGGGGGCGCGCGGGGGGGGGGGGGGGGGGGGGG	21963	
QY	636	AGCAGAGATTACAGCCGATTCATTAGGCAAAATGAGGCCCGCGCTGGGTGGGGGTGTGTG	695	
Db	21962	CCGGGGGCGGGCCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	21903	
QY	696	TTAAGGGAGGACACC GG	714	
Db	21902	GGG	21884	

```

RESULT 11
US-09-165-264-11/c
; Sequence 11, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoorthy, Thuraiyaiah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 11
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Primer sequence
US-09-165-264-11

```

[illegible][illegible]

```

RESULT 12
US-09-165-264-13/c
; Sequence 13, Application US/09165264
; Patent NO. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamoothy, Thuraiyavah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 13
; LENGTH: 320
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Primer sequence
US-09-165-264-13

```

	Query Match	4.8%;	Score 47.8;	DB 4;	Length 320;
	Best Local Similarity	47.8%;	Pred. No. 0.0038;		
	Matches 139;	Conservative 0;	Mismatches 152;	Indels 0;	Gaps
QY	710	CCGGGACACCGCCCTTTC	CCCCGCCACACACTCTC	CACAGGGTTTCGTGCGCA	76
Db	306	CCAGTGCCCCCCC	CCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCC	24
QY	770	GGGACTGACCAAACTTGG	GGGAGCCTGGGAGCGGA	CTGTTACAAGGGAGACGCC	82
Db	246	CCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCC	18
QY	830	GCCCTCTTCGTCCTTGT	TGCCTCGAGCCCCCTCT	CTCCTCGTACTCGGCTGCC	88
Db	186	CCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCC	12
QY	890	TGTACTGTGTACTCTCAT	CTGTGAGCCTTTC	CCCTTCCCTCTCTCTCTCT	94
Db	126	CCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCC	67
QY	950	CCCTTCCAGGCTGCC	CCCACTTGCCTGT	CTCACATGCGCGCTCTC	1000
Db	66	CCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCC	CCCCCCCCCCCCCCCC	16

```

RESULT 13
US-09-165-264-13
; Sequence 13, Application US/09165264
; Patent No. 6197510
; GENERAL INFORMATION:
; APPLICANT: Vinayagamorthy, Thuraiayah
; TITLE OF INVENTION: Multi-Loci Genomic Analysis
; FILE REFERENCE: 44747
; CURRENT APPLICATION NUMBER: US/09/165,264
; CURRENT FILING DATE: 1998-10-01
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn Ver. 2.1

```









Db 303 DSGSESSWSTGSSGNHGGGNGHKPCGCKPGRNARGSGESGICQFRQGVSSNM 362  
Qy 624 CCAGCGGAGTTACGAGAGATTACGCCGATTGCGATTAGG 663  
Db 363 EISKEGNRLGSGDNYRGQSSWGGGSDAVGGVNTVNS 402

## RESULT 7

US-09-954-456-292  
; Sequence 292, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using Candi  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 292  
; LENGTH: 43058  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-456-292

Query Match 4.8%; Score 48.4; DB 10; Length 43058;  
Best Local Similarity 52.4%; Pred. No. 0.00056;  
Matches 130; Conservative 0; Mismatches 116; Indels 2; Gaps 1;  
Qy 498 CGCGTCTCCGGGGAGGTGCGCGCGGCGAGG--GGGCTGGCGGGCGAGCCCC 555  
Db 25640 CGCGGGCGGGCGGGTTCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 25699  
Qy 556 CGCGGGGGGTGCGGAGCGGGTGATGTCAGGGCAGCGGTGGGTGCTACTCGGAGGTG 615  
Db 25700 GGGCCCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 25759  
Qy 616 AGCGCGCGCCAGCGAGTTCAGCGAGAGTTCAGCCGCGATTGCGATTAGGCAATGAGCCCC 675  
Db 25760 CGGGCGGGCGGGCGGGTTCGCGGGCGGGGTTCGCGGGCGGGCGGGCGGGCGG 25819  
Qy 676 GGCGTGGGTGGGGTGTGTGTTAAGGGAGGACACCGGGGACACCCCGCTTCCCGCGCC 735  
Db 25820 GGGCGGGGTGGGTGCGGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 25879  
Qy 736 CCACCACC 743  
Db 25880 CCGCCCCC 25887

## RESULT 8

US-09-954-456-529

; Sequence 529, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 529  
; LENGTH: 43058  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-456-529

Query Match 4.8%; Score 48.4; DB 10; Length 43058;  
Best Local Similarity 52.4%; Pred. No. 0.00056;  
Matches 130; Conservative 0; Mismatches 116; Indels 2; Gaps 1;  
Qy 498 CGCGTCTCCGGGGAGGTGCGCGCGGCGAGG--GGGCTGGCGGGCGAGCCCC 555  
Db 25640 CGCGGGCGGGCGGGTTCGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 25699  
Qy 556 CGCGGGGGGTGCGGAGCGGGTGATGTCAGGGCAGCGGTGGGTGCTACTCGGAGGTG 615  
Db 25700 GGGCCCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGG 25759  
Qy 616 AGCGCGCGCCAGCGAGTTCAGCGAGAGTTCAGCCGCGATTGCGATTAGGCAATGAGCCCC 675  
Db 25760 CGGGCGGGCGGGCGGGTTCGCGGGCGGGGTTCGCGGGCGGGCGGGCGGGCGG 25819  
Qy 676 GGCGTGGGTGGGGTGTGTGTTAAGGGAGGACACCGGGGACACCCCGCTTCCCGCGCC 735  
Db 25820 GGGCGGGGTGGGTGCGGGGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGCGG 25879  
Qy 736 CCACCACC 743  
Db 25880 CCGCCCCC 25887

## RESULT 9

US-09-880-107-3950  
; Sequence 3950, Application US/09880107  
; Patent No. US20020142981A1  
; GENERAL INFORMATION:  
; APPLICANT: Horne, Darci T.  
; APPLICANT: Vockley, Joseph G.  
; APPLICANT: Scherf, Uwe  
; APPLICANT: Gene Logic, Inc.  
; TITLE OF INVENTION: Gene Expression Profiles in Liver Cancer  
; FILE REFERENCE: 44921-5028-WO

```
; CURRENT APPLICATION NUMBER: US/09/880,107
; CURRENT FILING DATE: 2001-06-14
; PRIOR APPLICATION NUMBER: US 60/211,379
; PRIOR FILING DATE: 2000-06-14
; PRIOR APPLICATION NUMBER: US 60/237,054
; PRIOR FILING DATE: 2000-10-02
; NUMBER OF SEQ ID NOS: 3950
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3950
; LENGTH: 43058
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: Genbank Accession No. US20020142981A1 Z84721
US-09-880-107-3950

Query Match          4.8%; Score 48.4; DB 10; Length 43058;
Best Local Similarity 52.4%; Pred. No. 0.00056;
Matches 130; Conservative 0; Mismatches 116; Indels 2; Gaps 1;

QY 498 CCGCTCTCCGGGAGGTGCGCGGTTCCGCGGCGAGG--GGGCTGGCGGCGAGCCCC 555
Db 25640 CCGCGGCGCGCGGCGGCGGTCGCGGGGCGGCGGCGGCGGCGGCGGCGGCGCG 25699

QY 556 GCGGGCGGCGCTGCGGAGCGGTGATCTCACGGGCAGCGGTGGTGTCTACTCGGAGGTG 615
Db 25700 GGGCCCGCGGCGGCGGCGGCGGCGGAGGCGCTGGCGGGGCGGCGGCGGCGGCG 25759

QY 616 AGGCGCGCGCAGCGAGTTCAGCGAGTTCACCGCATTTAGGCAAAATCAGGCCCC 675
Db 25760 CCGGCGGCGCGGCGGCGGCGGTCGCGGGGCGGCGGCGGCGGCGGCGGCGGCG 25819

QY 676 GGCCTGGTGGGCTGTGTTAAGGGAGGACACCGGAGGACACCCCTCTCCCGGCC 735
Db 25820 GGGCGGGGTGGGCTGCGGGGCGGCGGCGGCTAGGCCCCCGCGCGACTGAGCGGC 25879

QY 736 CCACCACC 743
Db 25880 CCGGCCCC 25887

RESULT 10
US-10-095-407-16/c
; Sequence 16, Application US/10095407
; Patent No. US20020164330A1
; GENERAL INFORMATION:
; APPLICANT: Pan, Yang
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY
; FILE REFERENCE: 09404/052001
; CURRENT APPLICATION NUMBER: US/10/095,407
; CURRENT FILING DATE: 2002-03-11
; PRIOR APPLICATION NUMBER: US 60/091,650
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: US 60/054,646
; PRIOR FILING DATE: 1997-08-04
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 16
; LENGTH: 152331
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(152331)
; OTHER INFORMATION: n = A,T,C or G
US-10-095-407-16

Query Match          4.8%; Score 48.4; DB 9; Length 152331;
Best Local Similarity 51.7%; Pred. No. 0.00068;
Matches 134; Conservative 0; Mismatches 122; Indels 3; Gaps 1;

QY 456 GCGCCCGCGGACCGCGCCCTACGCTCTCGCGCGGCTCTCCGCGGAGGT 515
```

```
Db 22139 GGGCGGCGGCGGCGGCGTCTCGGCGGCGGCGGCGGCGGCGGCGTCTCGGCG 22080
QY 516 GGGCGGCGTTCGCGCGGCGGCGGCGTCTCGGCGGCGGCGGCGGCGGCGGCGGCG 575
Db 22079 GGTGCGTGTGGCGGCGGCGTGTGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 22020
QY 576 GTGATGTACGCGGCGGCGGCGGCGTGTGGGTCCTCGGAGGTGAGGCGGCCCGCGG 635
Db 22019 GTGGGGG---GCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21963
QY 636 AGCGAGAGTTTCAGCGCATTCATTAGGCAAAATGAGGCGCCGCGCTGGGTGGGGGTGTGTG 695
Db 21962 CGCGGGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 21903
QY 696 TTAAGGGGAGACACCGGG 714
Db 21902 GGGGGGGGGGGGGGGGG 21884

RESULT 11
US-09-771-208-20
; Sequence 20, Application US/09771208
; Patent No. US20020155564A1
; GENERAL INFORMATION:
; APPLICANT: MEDRANO, JUAN
; APPLICANT: BRADFORD, ERIC
; APPLICANT: HORVATH, SIMON
; TITLE OF INVENTION: CLONING OF A HIGH-GROWTH GENE
; FILE REFERENCE: 407T-923710US
; CURRENT APPLICATION NUMBER: US/09/771,208
; CURRENT FILING DATE: 2001-01-26
; PRIOR APPLICATION NUMBER: US 08/999,477
; PRIOR FILING DATE: 1997-12-29
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 20
; LENGTH: 659158
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (123459)..(123478)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (602466)..(602485)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (546998)..(547017)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (494715)..(494814)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (390986)..(391005)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (346860)..(346823)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (317174)..(317193)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (280353)..(280373)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (271829)..(271848)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (183872)..(183891)
; OTHER INFORMATION: n is unidentified a, c, g, or t
; NAME/KEY: misc_feature
; LOCATION: (170625)..(170645)
```





;; TITLE OF INVENTION: NOVEL LOW DENSITY LIPOPROTEIN BINDING  
;; TITLE OF INVENTION: PROTEINS AND THEIR USE IN DIAGNOSING AND TREATING  
;; TITLE OF INVENTION: ATHEROSCLEROSIS  
;; FILE REFERENCE: 10797-004001  
;; CURRENT APPLICATION NUMBER: US/10/023,523  
;; CURRENT FILING DATE: 2001-12-17  
;; PRIOR APPLICATION NUMBER: US/09/616,289  
;; PRIOR FILING DATE: 2000-07-14  
;; PRIOR APPLICATION NUMBER: US 09/517,849  
;; PRIOR FILING DATE: 2000-03-02  
;; PRIOR APPLICATION NUMBER: US 08/979,608  
;; PRIOR FILING DATE: 1997-11-26  
;; PRIOR APPLICATION NUMBER: US 60/031,930  
;; PRIOR FILING DATE: 1996-11-27  
;; PRIOR APPLICATION NUMBER: US 60/048,547  
;; PRIOR FILING DATE: 1997-06-03  
;; NUMBER OF SEQ ID NOS: 53  
;; SOFTWARE: FastSeq for Windows Version 4.0  
;; SEQ ID NO 48  
;; LENGTH: 2561  
;; TYPE: DNA  
;; ORGANISM: Oryctolagus cuniculus  
;; FEATURE:  
;; NAME/KEY: CDS  
;; LOCATION: (246)...(1895)  
US-10-023-523-48

Query Match 4.5%; Score 45; DB 12; Length 2561;  
Best Local Similarity 54.7%; Pred. No. 0.0032;  
Matches 110; Conservative 0; Mismatches 90; Indels 1; Gaps 1;  
Qy 505 CGGGGGAGGTGCGCGGTCGCGGCGAGGGGCTGCGGCGAGCCCGCGCGCGG 564  
Db 854 CGGGGGCG 795  
Qy 565 CTGGCGAGCGGTGATGTCACGGGAGCGGTGGTGGTCACTCGGAGGTGAGCGCCCG 624  
Db 794 CGGGCGCGCGCGAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 736  
Qy 625 CAGCGGAGTTCAGCGAGGTTCAGCGCGCATTCATTAGCAATAGCGCCCGCGCTGGGT 684  
Db 735 GCGAGGGGGCGCTGTGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 676  
Qy 685 GGGGTGTGTGTTAAGGGGAG 705  
Db 675 CGGGCGCGCGCGCGCGCGCG 655

RESULT 15  
US-10-032-393-47  
; Sequence 47, Application US/10032393  
; Publication No. US20030027286A1  
; GENERAL INFORMATION:  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Wall, Daniel  
; APPLICANT: Gross, Molly  
; TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE  
; FILE REFERENCE: ELITRA.010A  
; CURRENT APPLICATION NUMBER: US/10/032,393  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/259,434  
; PRIOR FILING DATE: 2000-12-27  
; PRIOR APPLICATION NUMBER: 09/948,993  
; PRIOR FILING DATE: 2001-09-06  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; NUMBER OF SEQ ID NOS: 68  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 47  
; LENGTH: 12733  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:

;; OTHER INFORMATION: Vector pEPEF14  
US-10-032-393-47  
Query Match 4.5%; Score 45; DB 9; Length 12733;  
Best Local Similarity 48.0%; Pred. No. 0.0042;  
Matches 129; Conservative 0; Mismatches 140; Indels 0; Gaps 0;  
Qy 507 GGGGGAGGTGCGCGGTCGCGGCGAGGGGCTGCGGCGAGCCCGCGCGCGGCT 566  
Db 5144 GGG 5203  
Qy 567 GCGAGCGGTGATGTCACGGGCGAGCGGTGGTGGTCACTCGGAGGTGAGCGCCCGCA 626  
Db 5204 GGG 5263  
Qy 627 GCGGAGTTCAGCGAGAGTTCAGCGCGCATTCATTAGCAATAGCGCCCGCGCTGGGTGG 686  
Db 5264 GGG 5323  
Qy 687 GGTGTGTGTTAAGGGGAGGACACCGGGGACACCCCGCTCTTCCCGCGCCACACCTCC 746  
Db 5324 GGG 5383  
Qy 747 TCCACCACGGCTTCGCTCGGCCAGGGACT 775  
Db 5384 GCTCTTCCGGTGGCGCGGGGGCATGACT 5412

Search completed: March 30, 2003, 16:45:13  
Job time : 956.893 secs

Result No.	Score	Query %		Length	DB	ID	Description
		Match					
C 1	656.2	65.6	929	14	BQ678320		BQ678320 AGENCOURT
C 2	590	59.0	602	14	BM702028		BM702028 UI-E-CQ1-
C 3	567.4	56.7	603	14	BM713229		BM713229 UI-E-E70-
C 4	563	56.3	591	14	BM726559		BM726559 UI-E-E70-
C 5	535	53.5	579	14	BQ931567		BQ931567 UI-E-E71-
C 6	517.6	51.8	652	14	BM981448		BM981448 UI-CF-BN1

LOCUS BM/020208 802 bp linear EST 28-FEB-2002

Qy	111	CTTTCTTCCAGTGAACGTCATCCCGAGGGCTCTTAGGATGAAGTAGTCCACTGGAAGG	170
Db	602	CTTTCTTCCAGTGAACGTCATCCCGAGGGCTTTTAGTGAAGTAGTCCACTGGAAGG	543
Qy	171	CACGAGCTCTTCCTTTTATCTCTTCACAGCTGGACAGTGCACAGGGGCCGTACTGGTT	230
Db	542	CACGAGCTTTCCTTTTATCTCTTCACAGCTGGACAGTGCACAGGGGCCGTACTGGTT	484
Qy	231	CCCCAGCTAGGAGACACCTTGGCGGGGGCTTTTGCTCCCGGAACGACGAGCCGTGGG	290

RESULT 2  
BM702028/c  
LOCUS

```
Db 483 CCCAGCTAGGAGACACCTTTGGCGGGCTTGTCTCCCGAAGACGACGAGCGTGGG 424
QY 291 AGAGGGCCCCCTCTGCTGTGTTGGCAACAGCACCGCGCTGCGGCTGGGTTC 350
Db 423 AGAGGGCCCCCTCTGCTGTGTTGGCAACAGCACCGCGCTGCGGCTGGGTTC 364
QY 351 GGGCGCGGAGTCACACATGATGTCACAGACAAATGACACAGCCGGTGTCTCATTCGGAC 410
Db 363 GGGCGCGGAGTCACACATGATGTCACAGACAAATGACACAGCCGGTGTCTCATTCGGAC 304
QY 411 ACAGGTCGCGAGTCACAAATGTCACACCCGGGTGCCAAACACTTGGCCCGCGACCC 470
Db 303 ACAGGTCGCGAGTCACAAATGTCACACCCGGGTGCCAAACACTTGGCCCGCGACCC 244
QY 471 GGCCTACGCTCTCTGCGCGGCTCTCCGGTCTCCGGGGAGGTGCGCGGTCGGCG 530
Db 243 GGCCTACGCTCTCTGCGCGGCTCTCCGGTCTCCGGGGAGGTGCGCGGTCGGCGG 184
QY 531 GGCAGGGGGTGGCGGCGAGCCCGCGGGCTGGCGAGCGGGTGTGTCACGGGCA 590
Db 183 GGCAGGGGGTGGCGGCGAGCCCGCGGGCTGGCGAGCGGGTGTGTCACGGGCA 124
QY 591 GCGGTGGGTGGTCACTCGGAGGTGAGCGCGCGGAGGTTACGCGAGTTCAGCC 650
Db 123 GCGGTGGGTGGTCACTCGGAGGTGAGCGCGCGGAGTTCAGGAGTTCAGCC 64
QY 651 GCATTCATTAGGCAATAGCGCGCGCTGGGTGGGTGGGTGGTGGTAAAGGGAGGAC 710
Db 63 GCATTCATTAGGCAATAGCGCGCGCTGGGTGGGTGGGTGGTGGTAAAGGGAGGAC 4
QY 711 CGG 713
Db 3 CGG 1

RESULT 3
BM713229/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLES
JOURNAL
MEDLINE
COMMENT

603 bp mRNA linear EST 28-FEB-2002
UI-E-EJ0-ahn-h-15-0-UI.r1 UI-E-EJ0 Homo sapiens cDNA clone
BM713229
BM713229
BM713229.1 GI:19026487
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 603)
Bonaldi, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
The following repetitive elements were found in this cDNA
sequence: 138-191, >(CGGG)n$imple_repeat (matched complement)
Seq primer: M13 Reverse.
Location/Qualifiers
1..603
/organism="Homo sapiens"
/db_xref="taxon:9606"
```

```
/clone="UI-E-EJ0-ahn-h-15-0-UI"
/clone_lib="UI-E-EJ0"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EJ0 is a subtracted cDNA library constructed
according to Bonaldi, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dr)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAAGA
; lens, CGATTAGCGA; eye anterior segment, AATGCCGCTAT;
optic nerve, CCATTAAGTG; retina, CCGCG; Retina Foveal and
Macular, GTCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye-Institute (NEI)."
```

BASE COUNT 105 a 203 c 199 g 96 t

ORIGIN

Query Match 56.7%; Score 567.4; DB 14; Length 603;  
Best Local Similarity 99.8%; Pred. No. 1.9e-120;  
Matches 568; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 113 TTCTTCCAGTGAACGTCATCCCGAGGGCTTCTAGGATGAAGTAGTCCACCTGGAAGGCA 172  
Db 603 TTCTTCCAGTGAACGTCATCCCGAGGGCTTCTAGGATGAAGTAGTCCACCTGGAAGGCA 544  
QY 173 CCAGCTCTTCTCTTATCTCTCCAGAGCTGGACAGTGACACAGGCGCGGTACTGTCTTC 232  
Db 543 CCAGCTCTTCTCTTATCTCTCCAGAGCTGGACAGTGACACAGGCGCGGTACTGTCTTC 484  
QY 233 CCAGCTAGGAGACACACTTTGGCGGGGCTTTGCTCCCGGAAGCAGCAGAGCGTGGGAG 292  
Db 483 CCAGCTAGGAGACACACTTTGGCGGGGCTTTGCTCCCGGAAGCAGCAGAGCGTGGGAG 424  
QY 293 GAGGCGCCCTCTGCTGTGTTGTGCCAACAGACACCGCGCTGCGCGGTTCGGTTCGG 352  
Db 423 GAGGCGCCCTCTGCTGTGTTGTGCCAACAGACACCGCGCTGCGCGGTTCGGTTCGG 364  
QY 353 CGGCGGAGTCACACATGATGTCACAGACAAATGACACAGCGCGGTGTCTCATTCGGACAC 412  
Db 363 CGGCGGAGTCACACATGATGTCACAGACAAATGACACAGCGCGGTGTCTCATTCGGACAC 304  
QY 413 AGCGTCGAGCTGCACAAATGTCACACCGGGGTGCCAAACACTTGGCCCGCGAGACCCGG 472  
Db 303 AGCGTCGAGCTGCACAAATGTCACACCGGGGTGCCAAACACTTGGCCCGCGAGACCCGG 244  
QY 473 CCCTAGCGCTCTGCGCGGCTCTCCGGGTCTCCGGGGAGGTGCGCGGTTCGGCGGG 532  
Db 243 CCCTAGCGCTCTGCGCGGCTCTCCGGGTCTCCGGGGAGGTGCGCGGTTCGGCGGG 184  
QY 533 CAGGGGCTGGCGGCGAGCCCGCGGGCTGGCGAGCGGGTGTGTCACGGGCGAC 592  
Db 183 CAGGGGCTGGCGGCGAGCCCGCGGGCTGGCGAGCGGGTGTGTCACGGGCGAC 124  
QY 593 GGTGGGTGGGTCACTCGGAGGTGAGCGCGCGGAGTTCAGCGAGAGTTCAGCCCG 652  
Db 123 GGTGGGTGGGTCACTCGGAGGTGAGCGCGCGGAGTTCAGCGAGAGTTCAGCCCG 64  
QY 653 ATTGCATTAGGCAAAATAGGCCCGCGCTG 681  
Db 63 ATTGCATTAGGCAAAATAGGCCCGCGCTG 35

RESULT 4



```
/db_xref="taxon:9606"
/clone="UI-E-EJ1-ajk-j-22-0-UI"
/clone_lib="UI-E-EJ1"
/tissue_type="fetal eyes, lens, eye anterior segment,
optic nerve, retina, Retina Foveal and Macular, RPE and
Choroid"
/dev_stage="fetal and adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/notes="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-E-EJ1 is a subcloned cDNA library constructed
according to Bonaldo, Lennon and Soares, Genome Research,
6:791-806, 1996. First strand cDNA synthesis was primed
with an oligo-dT primer containing a Not I site. Double
stranded cDNA was ligated to an EcoR I adaptor, digested
with Not I, and cloned directionally into pT73-Pac
vector. The oligonucleotide used to prime the synthesis of
first-strand cDNA contains a library tag sequence that is
located between the Not I site and the (dT)18 tail. The
sequence tags for this library are: fetal eyes, AGAATCAAGA
; lens, CGATTAGCGA; eye anterior segment, AATGCCGAT;
optic nerve, CCATTAGTGT; retina, CCGCG; Retina Foveal and
Macular, GPCC; RPE and Choroid, ACCTA. This library was
created for the program, Gene Discovery in the Visual
System, supported by National Eye Institute (NEI)."
BASE COUNT 100 a 197 c 186 g 95 t 1 others
ORIGIN
Query Match 53.5%; Score 535; DB 14; Length 579;
Best Local Similarity 99.5%; Pred. No. 5.5e-113;
Matches 568; Conservative 0; Mismatches 0; Indels 3; Gaps 3;

Qy 112 TTTCTTCAGTGAACGTCGATCCGAGGGCTTCTAGTAGTAAGTAGTCCACTGGAAGCG 171
Db 579 TTTCTTCAGTGAACGTCGATCCGAGGGCTTCTAGTAGTAAGTAGTCCACTGGAAGCG 520

Qy 172 ACCAGCTCTCTTTATCTCTC-CAGAGCTGGACAGTGCACAGGGGCGGTACTGGTT 230
Db 519 ACCAGCTCTCTTTATCTCTCNCAGAGCTGGACAGTGCACAGGGGCGGTACTGGTT 460

Qy 231 CCCAGCTAGGAGACACCTTGGCGGGGCTTGTCTCGCGGAGACGACGAGCGTGGGG 290
Db 459 CCCAGCTAGGAGACACCTTGGGC-GGGCTTGTCTCGCGGAGACGACGAGCGTGGGG 401

Qy 291 AGGAGGGCCCTCTGCTGCTGTTGTGCCAAGACACCCGCGTCCCGCTGCGGTTC 350
Db 400 AGGA-GGCGCCCTCTGCTGCTGTTGTGCCAAGACACCCGCGTCCCGCTGCGGTTC 342

Qy 351 GCGGCGGAGTCACACATGATGCACAGACATGACACAGCCGCTGCTCATTCGCGAC 410
Db 341 GCGGCGGAGTCACACATGATGCACAGACATGACACAGCCGCTGCTCATTCGCGAC 282

Qy 411 ACAGGCTCCGAGTCGACAAATGCACCCGGGTGCGCAACACTTGGCCCGCGGACCC 470
Db 281 ACAGGCTCCGAGTCGACAAATGCACCCGGGTGCGCAACACTTGGCCCGCGGACCC 222

Qy 471 GGCCTACGCTCTCTGCGCGCTCTCCGCGTCTCCGGGGAGGTGGCCCGGTTCGCGG 530
Db 221 GGCCTACGCTCTCTGCGCGCTCTCCGCGTCTCCGGGGAGGTGGCCCGGTTCGCGG 162

Qy 531 GGCAGGGGCTGGCGGCGAGCCCGCGGGGCTGGCGAGCGGTGATGCAGGGCA 590
Db 161 GGCAGGGGCTGGCGGCGAGCCCGCGGGGCTGGCGAGCGGTGATGCAGGGCA 102

Qy 591 GCGGTGGTGGTCACTCGGAGGTGAGCGCGCGGAGGTTCAGCGAGAGTTCAGCC 650
Db 101 GCGGTGGTGGTCACTCGGAGGTGAGCGCGCGGAGGTTCAGCGAGAGTTCAGCC 42

Qy 651 GCATTGCAATAGGCAATGAGCCCGGCGT 681
Db 41 GCATTGCAATAGGCAATGAGCCCGGCGT 11
```

```
RESULT 6
BM981448
LOCUS
DEFINITION
  BM981448 652 bp mRNA linear EST 21-MAR-2002
  UI-CF-EN1-adh-p-17-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone
  UI-CF-EN1-adh-p-17-0-UI 3', mRNA sequence.
ACCESSION
  BM981448
VERSION
  BM981448.1 GI:19603947
KEYWORDS
  EST.
SOURCE
  human.
ORGANISM
  Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
  1 (bases 1 to 652)
  Bonaldo,M.F., Lennon,G. and Soares,M.B.
  Normalization and subtraction: two approaches to facilitate gene
  discovery
  Genome Res. 6 (9), 791-806 (1996)
JOURNAL
  97044477
MEDLINE
  97044477
COMMENT
  Contact: McCray, PB
  McCray Lab
  University of Iowa
  2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
  Tel: 319 356 4866
  Fax: 319 356 7171
  Email: paul-mccray@uiowa.edu
  Tissue Procurement: Dr. M. J. Welsh, University of Iowa
  cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
  cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
  DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
  Clone Distribution: Researchers may obtain clones from Research
  Genetics (www.resgen.com).
  Seq primer: M13 FORWARD
  POLYA=Yes.
FEATURES
  Location/Qualifiers
    1..652
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /clone="UI-CF-EN1-adh-p-17-0-UI"
      /clone_lib="UI-CF-EN1"
      /tissue_type="Primary Lung Cystic Fibrosis Epithelial
      Cells"
      /dev_stage="Adult"
      /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
      /notes="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
      modified polylinker; Site_1: EcoR I; Site_2: Not I;
      UI-CF-EN1 is a normalized cDNA library containing the
      following tissue(s): Primary Lung Cystic Fibrosis
      Epithelial Cells. The library was constructed according to
      Bonaldo, Lennon and Soares, Genome Research, 6:791-806,
      1996. First strand cDNA synthesis was primed with an
      oligo-dT primer containing a Not I site. Double stranded
      cDNA was ligated to an EcoR I adaptor, digested with Not
      I, and cloned directionally into pT73-Pac vector. The
      oligonucleotide used to prime the synthesis of
      first-strand cDNA contains a library tag sequence that is
      located between the Not I site and the (dT)18 tail. The
      sequence tag for this library is CTGCTCAGGT.
      TAG_LTB=UI-CF-EN1
      TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
      6hr to LPS 24h
      TAG_SEQ=CTGCTCAGGT"
BASE COUNT 124 a 192 c 177 g 158 t 1 others
ORIGIN
Query Match 51.8%; Score 517.6; DB 14; Length 652;
Best Local Similarity 98.1%; Pred. No. 5.8e-109;
Matches 523; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

Qy 1 GTCCCTGGGTAGCATGTACATTTCCATCCCTCTCTTTATATATATGGGGTAATAGGATAC 60
Db 120 GTCCCTGGGTAGCATGTACATTTCCATCCCTCTCTTTAGAGAGTGGGGTAATAGGATAC 179
Qy 61 CCCCTCCTCCAGGGGTATCCCTCTTTCTAGGAGACCTACCCAAAGCTAGGCTTCTTCCCA 120
```



```
Db 180 CCCTCTCCAGGGGTATCCCTCTTCTAGGACCTACCAAGCTAGGCGCTTCTTCCA 239
Qy 121 GTGAAGCTGCATCCCGAGGGCTCTAGGATGAAGTAGTCCACTGGAGGCACCAAGCTCT 180
Db 240 GTGAAGCTGCATCCCGAGGGCTCTAGGATGAAGTAGTCCACTGGAGGCACCAAGCTCT 299
Qy 181 TCCTTTTATCTCTCCAGAGCTGGACAGTGCACAGGGCGGGTACTGGTTCCTCCAGTAG 240
Db 300 TCCTTTTATCTCTCCAGAGCTGGACAGTGCACAGGGCGGGTACTGGTTCCTCCAGTAG 359
Qy 241 GAGACACTTGGGGGGGGCTTGTGTCGCCGAAGCAGCAGAGCGTGGGAGGAGGCC 300
Db 360 GAGACACTTGGGGGGGGCTTGTGTCGCCGAAGCAGCAGAGCGTGGGAGGAGGCC 419
Qy 301 CCTCTGCTGTGTTGTGCCAAGCAGCAGCGCGTGCCTGGGTTCCGGCGCGGA 360
Db 420 CCTCTGCTGTGTTGTGCCAAGCAGCAGCGCGTGCCTGGGTTCCGGCGCGGA 479
Qy 361 GTCACATGATGTCACAGACATGACACAGCGGTGTCTCATTCGACACAGCGTCCG 420
Db 480 GTCACATGATGTCACAGACATGACACAGCGGTGTCTCATTCGACACAGCGTCCG 539
Qy 421 AGCTGCACATGTCACACCGGGTGCACAACTTGGCCCGCGGACCGCGCTTACGC 480
Db 540 AGCTGCACATGTCACACCGGGTGCACAACTTGGCCCGCGGACCGCGCTTACGC 599
Qy 481 CTCTCGCGCGCTCTCCGGTCTCCGGGGAGGTGCGCGGTTCGGCGCGGC 533
Db 600 CTCTCGCGCGCTCTCCGGTCTCCGGGGAGGTGCGCGGTTCGGCGCGGC 652

RESULT 7
BM684023
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
COMMENT
```

FEATURES

Source  
1. 627  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="UI-E-EJ1-ajk-j-22-0-UI"  
/clone\_lib="UI-E-EJ1"  
/tissue\_type="fetal eyes, lens, eye anterior segment, optic nerve, retina, Retina Foveal and Macular, RPE and

Choroid"  
/dev\_stage="fetal and adult"  
/lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
/note="Organ: eye; Vector: p7T3-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; UI-E-EJ1 is a subtracted cDNA library constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tags for this library are: fetal eyes, AGAATCAAGA; lens, CGATTAGCGA; eye anterior segment, AATGCGCGAT; optic nerve, CCATTAAGTG; retina, CCGCG; Retina Foveal and Macular, GTCC; RPE and Choroid, ACCTA. This library was created for the program, Gene Discovery in the Visual System, supported by National Eye Institute (NEI).  
TAG\_LIB=UI-E-EJ1  
TAG\_TISSUE=RPE and Choroid  
TAG\_SEQ=ACCTA"

BASE COUNT 123 a 185 c 164 g 155 t  
ORIGIN  
Query Match 49.5%; Score 494.6; DB 14; Length 627;  
Best Local Similarity 98.2%; Pred. No. 1.2e-103;  
Matches 500; Conservative 0; Mismatches 9; Indels 0; Gaps 0;  
Qy 1 GTCCTTGGGTAGCATGATTCATCCCTCTCTTTATATATGCGGGTATAGATAC 60  
Db 119 GTCCTTGGGTAGCATGATTCATCCCTCTTTAGAGAGTGGGGTATAGATAC 178  
Qy 61 CCCCTCTCCAGGGGTATCCCTCTTTAGGGAGCTACCCAAAGCTAGGCTTTCTTCCA 120  
Db 179 CCCCTCTCCAGGGGTATCCCTCTTTAGGGAGCTACCCAAAGCTAGGCTTTCTTCCA 238  
Qy 121 GTGAAGCTGCATCCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGGCACCAAGCTCT 180  
Db 239 GTGAAGCTGCATCCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGGCACCAAGCTCT 298  
Qy 181 TCCTTTTATCTCTCCAGAGCTGGACAGTGCACAGGGCGCGTACTGTTCCCTCCAGCTAG 240  
Db 299 TCCTTTTATCTCTCCAGAGCTGGACAGTGCACAGGGCGCGTACTGTTCCCTCCAGCTAG 358  
Qy 241 GAGACACTTGGGGGGGGCTTGTGTCGCCGAAGCAGCAGAGCGTGGGGAGGAGGCC 300  
Db 359 GAGACACTTGGGGGGGGCTTGTGTCGCCGAAGCAGCAGAGCGTGGGGAGGAGGCC 418  
Qy 301 CCTCTGCTGTGTTGTGCCAAGCAGCAGCGGTGCGCGGTTCGGGTTCGGCGCGCGGA 360  
Db 419 CCTCTGCTGTGTTGTGCCAAGCAGCAGCGGTGCGCGGTTCGGGTTCGGCGCGCGGA 478  
Qy 361 GTCACATGATGTCACAGCAATGACACAGCGGTCTCATTCGACACAGCGTCCG 420  
Db 479 GTCACATGATGTCACAGCAATGACACAGCGGTCTCATTCGACACAGCGTCCG 538  
Qy 421 AGCTGCACATGTCACACCGGGTGCACAACTTGGCCCGCGGACCGCGCTTACGC 480  
Db 539 AGCTGCACATGTCACACCGGGTGCACAACTTGGCCCGCGGACCGCGCTTACGC 598  
Qy 481 CTCTCGCGCGCTCTCCGGTCTCCGGGTTCGGCG 509  
Db 599 CTCTCGCGCGCTCTCCGGTCTCCGGGTTCGGCG 627

RESULT 8  
BM675116  
LOCUS  
DEFINITION  
ACCESSION  
631 bp mRNA linear EST 27-FEB-2002  
UI-E-EJ0-ahn-h-16-0-UI.s1 UI-E-EJ0 Homo sapiens cDNA clone  
UI-E-EJ0-ahn-h-16-0-UI 3', mRNA sequence.  
BM675116



```

VERSION BM675116.1 GI:18985014
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheraia; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 631)
NORMALIZATION and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 9704477
COMMENT Contact: Soares, MB
Program for Rat Gene Discovery and Mapping
University of Iowa
451 Eckstein Medical Research Building Iowa City, IA 52242, USA
Tel: 319 335 8250
Fax: 319 335 9565
Email: msoares@blue.weeg.uiowa.edu
Tissue Procurement: Dr. Gregg Hageman
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com).
Seq primer: M13 Forward
POLrA-Yes.

FEATURES
    Location/Qualifiers
        1..631
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="UI-E-EJ0-ahn-h-16-0-UI"
            /clone_lib="UI-E-EJ0"
            /tissue_type="fetal eyes, lens, eye anterior segment,
            optic nerve, retina, Retina Foveal and Macular, RPE and
            Choroid"
            /dev_stages="fetal and adult"
            /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
            /note="Organ: eye; Vector: pT73-Pac (Pharmacia) with a
            modified polylinker; Site_1: EcoR I; Site_2: Not I;
            UI-E-EJ0 is a subcloned cDNA library constructed
            according to Bonaldo, Lennon and Soares, Genome Research,
            6:791-806, 1996. First strand cDNA synthesis was primed
            with an oligo-dT primer containing a Not I site. Double
            stranded cDNA was ligated to an EcoR I adaptor, digested
            with Not I, and cloned directionally into pT73-Pac
            vector. The oligonucleotide used to prime the synthesis of
            first-strand cDNA contains a library tag sequence that is
            located between the Not I site and the (dnp)18 tail. The
            sequence tags for this library are: fetal eyes, AGAATCAGA
            ; lens, CGATTAGCGA; eye anterior segment, AATGCCGCAT;
            optic nerve, CCAATTAAGTG; retina, CCGCG; Retina Foveal and
            Macular, GTCC; RPE and Choroid, ACCTA. This library was
            created for the program, Gene Discovery in the Visual
            System, supported by National Eye Institute (NEI).
            TAG_LIB=UI-E-EJ0
            TAG_TISSUE=human fetal eyes
            TAG_SEQ=AGAATCAAGA"
BASE COUNT 124 a 184 c 168 g 155 t
ORIGIN

Query Match 48.5%; Score 485; DB 14; Length 631;
Best Local Similarity 97.9%; Pred. No. 1.9e-101;
Matches 502; Conservative 0; Mismatches 10; Indels 1; Gaps 1;

QY 1 GTCCTTGGGTACATGATCCATCCCTTCCTTTATATATATATAGGATAC 60
DB 120 GTCCCTGGGTACATGATCCATCCCTTCCTTTAGAGAGTGGGGTAATAGGATAC 179
QY 61 CCCCTCCCTCAGGGGTATCCCTCTTCTAGGAGCTACCCAGCTAGGCCCTTCTTCCA 120
DB 180 CCCCTCCCTCAGGGGTATCCCTCTTCTAGGAGCTACCCAGCTAGGCCCTTCTTCCA 239

```

```

QY 121 GTGAACGTCATCCCGAGGCGCTTCTAGGATGAAGTACTCCACTGGAAGGACCACTCT 180
DB 240 GTGAACGTCATCCCGAGGCGCTTCTAGGATGAAGTACTCCACTGGAAGGACCACTCT 299
QY 181 TCCTTTTATCTCTCCAGAGTGGACAGTGCACACAGGGGCGGTACTGTTCCCACTAG 240
DB 300 TCCTTTTATCTCTCCAGAGTGGACAGTGCACACAGGGGCGGTACTGTTCCCACTAG 359
QY 241 GAGACACCTTTGGGGGGGCTTTGCTCGCCGGAACACAGAGCGTGGGAGGAGGCC 300
DB 360 GAGACACCTTTGGGGGGGCTTTGCTCGCCGGAACACAGAGCGTGGGAGGAGGCC 418
QY 301 CCTCTGCTGTTGTGCGCAACAGACCCGCGTCCGCGTCCGCGTCCGCGCGCGCA 360
DB 419 CCTCTGCTGTTGTGCGCAACAGACCCGCGTCCGCGTCCGCGTCCGCGCGCGCA 478
QY 361 GTCACACATGATGTCACAGACAAATGACACAAAGCGGTGTCATTCACACAGCGTC 420
DB 479 GTCACACATGATGTCACAGACAAATGACACAAAGCGGTGTCATTCACACAGCGTC 538
QY 421 AGCTGCAATATGTCACACCGGGTGCCAAACACATTGGCCCGCGAGCCCGCCCTACGC 480
DB 539 AGCTGCAATATGTCACACCGGGTGCCAAACACATTGGCCCGCGAGCCCGCCCTACGC 598
QY 481 CTCTGCGCGCGCTCTCCGCGTCTCCGGGGGAG 513
DB 599 CTCTGCGCGCGCTCTCCGCGTCTCCGGGGGAG 631

RESULT 9
BI759786/c
LOCUS BI759786
DEFINITION 60304568F1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5185961 5',
mRNA sequence.
ACCESSION BI759786
VERSION BI759786.1 GI:15751364
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheraia; Primates; Catarrhini; Hominidae; Homo.
TITLE 1 (bases 1 to 797)
JOURNAL NIH-MGC http://mgc.nci.nih.gov/
COMMENT National Institutes of Health, Mammalian Gene Collection (MGC)
Contact: Robert Strausberg, Ph.D.
Email: cyabps-r@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)
Clone Distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLML1464 row: g column: 18
High quality sequence stop: 582.
Location/Qualifiers
    1..797
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="IMAGE:5185961"
        /clone_lib="NIH_MGC_116"
        /lab_host="DH10B"
        /note="Organ: pooled colon, kidney, stomach; Vector:
        PCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA
        source anonymous pool of 3 colons, age 26 yo male, 49 yo
        female, 71 yo male colon; 46 yo male kidney, and pool of 2
        stomachs, 62 yo male and 70 yo female. Library is
        oligo-dT primed and directionally cloned (EcoRV site is
        destroyed upon cloning). Average insert size 1.4 kb,
        insert size range 1-3 kb. Library is normalized and
        enriched for full-length clones and was constructed by C.
        Gruber (Invitrogen). Research Genetics tracking code

```

023. Note: this is a NIH\_MGC Library."

```
BASE COUNT      140 a  280 c  234 g  143 t
ORIGIN

Query Match      48.28; Score 482; DB 13; Length 797;
Best Local Similarity 92.5%; Pred. No. 1e-100;
Matches 618; Conservative 0; Mismatches 35; Indels 15; Gaps 10;

QY  29 CTTCTCTTTTATATATGGGGTATAGGATACCC-CTCCCTCCAGGGGTATCCCTC- 84
DB  596 CTTGTAGAGAGTGGGGGTATAGGATACCCGGTGGCTGAGGGGTATCGCTCGT 637

QY  85 TTTCTAGGACCTACCAAGCTAGGCTTTCTTCCAGTG-----AAACGTGCATCCCGAGG 140
DB  636 GTGCTAGGACCTAGCCGAGCTAGGCGCTTCCGTTCACAGGTGAACCCCTGCATCCCGAGG 577

QY  141 G-CTTCTAGGATGAGT-AGTCCACTGGAAGGCACCAGCTTCTCC-TTTTATCTCTCAG 197
DB  576 GCCTTCTAGGATGAGTAAAGTCCACTGGAAGGCACCAGCTTCTCCGGGTATCTCTCCAG 517

QY  198 AGCTGGACAGTGCACCAAGGCGCGTA-CTGGTTCCTCCAGCTAGGAGACACTTGGCGGG 256
DB  516 AGCTGGACAGTGCACCAAGGCGCGTACCTGTTCCAGCTAGGAGACACTTGGCGGG 457

QY  257 GG-CTTTGCTGCGCGGAAGCAGCAGAGCGTGGGAGAGGGCCCCCTC-TGCCCTGTG-T 313
DB  456 GGCCTTAGCTCGCGGAAGCAGCAGAGCGTGGGAGAGGGCCCCCTCGTGCCTGTGTT 397

QY  314 TTGTGCCAACAGCACCGCGCTGCGGTTCGGGTTCGGGCGCGGAGTGCACATGATG 373
DB  396 TTGTGCCAACAGCACCGCGCTGCGGTTCGGGTTCGGGCGCGGAGTGCACATGATG 337

QY  374 TCACAGACAATGACACAGCGGTGCTCATTCGACACAGCGTCCGAGCTGCACAATGT 433
DB  336 TCACAGACAATGACACAGCGGTGCTCATTCGACACAGCGTCCGAGCTGCACAATGT 277

QY  434 CACACCGGGTGCCAAACACTTGGCCCGCGCGAGCGCCCTAGCGCTCCTGCGCGCGC 493
DB  276 CACACCGGGTGCCAAACACTTGGCCCGCGCGAGCGCCCTAGCGCTCCTGCGCGCGC 217

QY  494 TCTCCGGCTCTCCGGGAGTGGCCGGTTCGGCGGGCAGGGGCTGGCGGGGAGCC 553
DB  216 TCTCCGGCTCTCCGGGAGTGGCCGGTTCGGCGGGCAGGGGCTGGCGGGGAGCC 157

QY  554 CCGCGGGCGGCTGCGGAGCGGTGATGTCACGGGACGCGTGGTGGGTCACTCGGAGG 613
DB  156 CCGCGGGCGGCTGCGGAGCGGTGATGTCACGGGACGCGTGGTGGGTCACTCGGAGG 97

QY  614 TGAGGCGCGCGAGCGGAGTTACGAGAGTTCAGCGGCATTCGATTAGGCAATGAGGC 673
DB  96 TGAGGCGCGCGAGCGGAGTTACGAGAGTTCAGCGGCATTCGATTAGGCAATGAGGC 37

QY  674 CCGGCCCTG 681
DB  36 CCGGCCCTG 29

RESULT 10
BQ011927
LOCUS      562 bp mRNA linear EST 26-MAR-2002
DEFINITION UI-1-BC1p-ati-b-01-0-UI.s1 NCI_CGAP_P13 Homo sapiens cDNA clone
            UI-1-BC1p-ati-b-01-0-UI 3', mRNA sequence.
ACCESSION  BQ011927
VERSION    BQ011927
KEYWORDS   EST.
-SOURCE   human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 562)
AUTHORS   NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE     National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
```

# JOURNAL COMMENT

Unpublished (1997)

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Dr. Steven Brown

cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa

Clone Distribution: Clone distribution information can be obtained  
from Dr. M. Bento Soares, bento-soares@uiowa.edu

Seq primer: M13 FORWARD

POLYA-yes.

Location/Qualifiers

1. 562

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="UI-1-BC1p-ati-b-01-0-UI"

/clone\_lib="NCI\_CGAP\_P13"

/tissue\_type="Placenta"

/dev\_stage="8-9 weeks"

/lab\_host="DH10B (Life Technologies)"

/note="Organ: Placenta; Vector: pT7T3-Pac (Pharmacia) with  
a modified polylinker; Site: 1: EcoR I; Site 2: Not I;

NCI\_CGAP\_P13 is a subtracted cDNA library constructed  
according to Bonaldo, Lennon and Soares, Genome Research,

6:791-806, 1996. First strand cDNA synthesis was primed  
with an oligo-dT primer containing a Not I site. Double

stranded cDNA was ligated to an EcoR I adaptor, digested  
with Not I, and cloned directionally into pT7T3-Pac

vector. The oligonucleotide used to prime the synthesis of  
first-strand cDNA contains a library tag sequence that is

located between the Not I site and the (dT)18 tail. The  
sequence tags for this library are GA, AGAA. For

additional information, contact: Bento Soares,

bento-soares@uiowa.edu

TAG\_LIB=UI-1-BC1p

TAG\_TISSUE=placenta human 8 week

TAG\_SEQ=GA

BASE COUNT 118 a 151 c 147 g 146 t

# ORIGIN

```
Query Match      42.68; Score 426.4; DB 14; Length 562;
Best Local Similarity 97.5%; Pred. No. 5.8e-88;
Matches 433; Conservative 0; Mismatches 11; Indels 0; Gaps 0;

QY  1 GTCCCTGGGTAGCATGATTTCCATCTCCCTCTTTATATATATATAGGGGTAAATAGGATAC 60
DB  119 GTCCCTGGGTAGCATGATTTCCATCTCCCTTTTAGAGAGTGGGGTAAATAGGATAC 178

QY  61 CCCCTCCTCCAGGGGTATCCCTCTTTCTAGGAGACCTACCCAGCTAGAGCCCTTTCTTCCA 120
DB  179 CCCCTCCTCCAGGGGTATCCCTCTTTCTAGGAGACCTACCCAGCTAGAGCCCTTTCTTCCA 238

QY  121 GTGAAACGTGCATCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGGCACCACTCT 180
DB  239 GTGAAACGTGCATCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGGCACCACTCT 298

QY  181 TCCTTTTATCTCTCCAGAGCTGGACAGTGCACAGGGCGGCTACTGTTTCCACAGCTAG 240
DB  299 TCCTTTTATCTCTCCAGAGCTGGACAGTGCACAGGGCGGCTACTGTTTCCACAGCTAG 358

QY  241 GAGACACCTTGGCGGGGCTTTGCTCGCGGGAAGCACGACGAGCGTGGGAGGAGGCGCC 300
DB  359 GAGACACCTTGGCGGGGCTTTGCTCGCGGGAAGCACGACGAGCGTGGGAGGAGGCGCC 418

QY  301 CCTCTGCTGTGTTTGTGCCAACAGACACCCGCGTGCCTCGCTGGGTTCGCGGGCGCGGA 360
DB  419 CCTCTGCTGTGTTTGTGCCAACAGACACCCGCGTGCCTCGCTGGGTTCGCGGGCGCGGA 478

QY  361 GTCACACATGATGTACACACATACACAAAGCCGGGTGTCATCTCCGACACAGCTCCG 420
DB  479 GTCACACATGATGTACACACATACACAAAGCCGGGTGTCATCTCCGACACAGCTCCG 538

QY  421 AGCTGCACAATGTCACACCCCGGT 444
```





QY	301	CCTCTGCCTGTGTTGTGCCAACAGACACCCCGCGTCCGCGTCCGGGTTC	CGGCGGCGG	360
Db	407			
Db	407	CCTCTGCCTGTGTTGTGCCAACAGACACCCCGCGTCCGCGTTC	CGGCGGCGG	466
QY	361	GTCCACATGATGTCACAGACAATGACACACAGCCGGTGCTCTCATTC	CCGACACAGGTC	420
Db	467	GTCCACATGATGTCACAGACAATGACACACAGCCGGTGCTCTCATTC	CCGACACAGGTC	526
QY	421	AGCT 424		
Db	527	AGCT 530		
RESULT	14			
BM849644/c				
LOCUS	BM849644	446 bp	mrna	linear
DEFINITION	K-EST0129836 S21SN520 Homo sapiens cdna clone S21SN520-36-A07 5'			EST 06-MAR-2002
ACCESSION	BM849644			
VERSION	BM849644.1	GI:19206043		
KEYWORDS	EST.			
SOURCE	human.			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
	1 (bases 1 to 446)			
	Kim,N.S., Chaeng,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,			
	Oh,K.J., Sohn,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and			
	Kim,Y.S.			
TITLE	21C Frontier Korean EST Project 2001			
JOURNAL	Unpublished (2002)			
COMMENT	Contact: Kim YS			
	Genome Research Center			
	Korea Research Institute of Bioscience & Biotechnology			
	52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea			
	Tel: +82-42-860-4470			
	Fax: +82-42-860-4409			
	Email: yongsung@mail.kribb.re.kr			
	Plate: 36 row: A column: 07			
	High quality sequence stop: 446.			
FEATURES	Location/Qualifiers			
source	1..446			
	/organism="Homo sapiens"			
	/db_xref="taxon:9606"			
	/clone="S21SN520-36-A07"			
	/clone_lib="S21SN520"			
	/sex="F"			
	/tissue_type="Stomach"			
	/cell_type="Floating aggregates"			
	/cell_line="SNU-520"			
	/lab_host="Top10F"			
	/notes="Organ: Stomach; Vector: pTZ189pl; Site.1: EcoRI;			
	Site.2: NotI; The poly (A)+ RNA was dephosphorylated with			
	bacterial alkaline phosphatase (BAP) and then decapped			
	with tabacco acid pyrophosphatase (TAP). The decapped			
	intact mRNA was ligated with DNA-RNA linker including EcoR			
	I site by treatment of T4 RNA ligase and the first strand			
	cdna was synthesized from oligo dT-selected mRNA by as			
	priming with dT-tailed vector. The dT-tailed vector was			
	adjusted to have about 60nt. The cdna vector was			
	circularized with E. coli DNA ligase after digestion of			
	EcoRI which site is also included in vector. An RNA strand			
	converted to a DNA strand by Okayama-Berg method. The			
	obtained cdna vectors were used for transformation of			
	competent cells E. coli Top10F by electroporation method.			
	The cdna libraries constructed by this method are			
	full-length enriched cdna library."			
BASE COUNT	73 a	156 c	151 g	65 t
ORIGIN				1 others
Query Match	40.6%	Score 406;	DB 14;	Length 446;
Best Local Similarity	99.8%	Pred. No. 2,7e-83;		

Double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT7T3 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT	117 a	148 c	143 g	134 t	1 others
ORIGIN					
Query Match	39.4%	Score	393.8	DB 9	Length 543
Best Local Similarity	96.6%	Pred No.	1.9e-80		
Matches	423	Conservative	0	Mismatches	13
				Indels	2
				Gaps	2
QY	1	GTCCCTGGGTAGCATGATACATTCCATTCCTTCCTTTATATATGGGGTAAATAGGATAC	60		
DB	107	GTCCCTGGGTAGCATGATACATTCCATTCCTTCCTTTAGAGAGTGGGGTAAATAGGATAC	166		
QY	61	CCCCTCCTCCAGGGTATCCCTCTTCTTAGGGACCTACCCAGCTAGGCGCTTCTTCCA	120		
DB	167	CCCCTCCTCCAGGGTATCCCTCTTCTTAGGGACCTACCCAGCTAGGCGCTTCTTCCA	226		
QY	121	GTGAACGTGTCATCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGSCACAGCTCT	180		
DB	227	GTGAACGTGTCATCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAAGSCACAGCTCT	286		
QY	181	TCCTTTTATCTCTCCAGAGCTGGACAGTGCACACAGGGCGCGGTACTGTGTTCCCCAGCTAG	240		
DB	287	TCCTTTTATCTCTCCAGAGCTGGACAGTGCACACAGGGCGCGGTACTGTGTTCCCCAGCTAG	346		
QY	241	GAGACACCTTGGGGGGGCTTTGCTGCGCGGAAGCACAGAGCGTGGGGA-GGAGGGCC	299		
DB	347	GAGACACCTTGGGGGGGCTTTGCTGCGCGGAAGCACAGAGCGTGGGGA-GGAGGGCC	406		
QY	300	CCCTCTGCTGTGTTGTGCCAACAGACACCGCGCTGCGCGTCCGGCGGCGCGG	359		
DB	407	CCCTCTGCTGTGTTGTGCCAACAGACACCGCGCTGCGCGTCCGGCGGCGCGG	465		
QY	360	AGTCACACATGATGCACAGACAATGACACAGCCGGTGTCTCATTTCCGACACAGCGTCC	419		
DB	466	AGTCACACATGATGCACAGACAATGACACAGCCGGTGTCTCATTTCCGACACAGCGTCC	525		
QY	420	GAGCTGCACAATGTCACA	437		
DB	526	GAGCTGCACAATGTCACA	543		

Search completed: March 30, 2003, 13:48:06  
Job time : 1296.28 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:00:38 ; Search time 2122.53 seconds  
(without alignments)  
13725.086 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_5000\_6000  
Perfect score: 1001  
Sequence: 1 tgatcccccctccccccacc.....tgggtcacccctcctggctg 1001

Scoring table: IDENTITY\_NUC  
Gap 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl :

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.ov.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.in.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pla.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	1001	100.0	16913	9	HSA297538	AJ297538 Homo sapi
2	1001	100.0	138999	9	AC090426	AC090426 Homo sapi
3	1001	100.0	158766	2	AC015851	AC015851 Homo sapi
4	1001	100.0	166368	2	AC018629	AC018629 Homo sapi
5	1001	100.0	173441	2	AC080112	AC080112 Homo sapi
6	1001	100.0	188574	2	AC131063	AC131063 Homo sapi
7	1001	100.0	190309	2	AC126392	AC126392 Homo sapi
8	584.2	58.4	13508	6	AX344790	AX344790 Sequence
9	507.2	50.7	13508	6	AX344791	AX344791 Sequence
10	384.2	38.4	516	11	HUMUT5218	L17737 Human STS U
11	146.4	14.6	162393	2	AC111746	AC111746 Rattus no
12	140.4	14.0	231248	2	AL591067	AL591067 Mus muscu
13	60	6.0	914	9	S57794	S57794 RARA=retino
14	59.2	5.9	125020	9	AF429315	AF429315 Homo sapi
15	55	5.5	194929	2	AC119084	AC119084 Rattus no
16	52.2	5.2	173898	2	AC109005	AC109005 Rattus no
17	52	5.2	7218	6	I66494	I66494 Sequence 14
18	51.6	5.2	164520	2	AC020738	AC020738 Homo sapi
19	51	5.1	174031	2	AC118303	AC118303 Rattus no
20	50	5.0	61633	2	AC084075	AC084075 Homo sapi
21	49	4.9	121473	2	AC099003	AC099003 Rattus no
22	48.6	4.9	58928	2	AC096234	AC096234 Rattus no
23	48.4	4.8	26822	2	AC011674	AC011674 Homo sapi
24	47.2	4.7	64055	2	AC111607	AC111607 Rattus no
25	46.8	4.7	171204	2	AC130936	AC130936 Rattus no
26	46.6	4.7	85268	2	AC022648	AC022648 Homo sapi
27	46.6	4.7	117082	2	AC126076	AC126076 Rattus no
28	46.6	4.7	165988	2	AC121212	AC121212 Rattus no
29	46.6	4.7	191277	2	AC128784	AC128784 Rattus no
30	46.4	4.6	130899	2	AC126730	AC126730 Rattus no
31	46.4	4.6	236562	2	AL772338	AL772338 Mus muscu
32	46.2	4.6	168280	2	AC123357	AC123357 Rattus no
33	46	4.6	101509	2	AC027353	AC027353 Homo sapi
34	46	4.6	125020	9	AF429315	AF429315 Homo sapi
35	45.8	4.6	138467	2	AC111697	AC111697 Rattus no
36	45.8	4.6	180758	2	AC128063	AC128063 Rattus no
37	45.8	4.6	196320	9	AL731559	AL731559 Human DNA
38	45.6	4.6	174845	2	AC128072	AC128072 Rattus no
39	45.4	4.5	188133	2	AC112428	AC112428 Rattus no
40	45.2	4.5	68330	2	AC116109	AC116109 Mus muscu
41	45.2	4.5	194776	2	AC129706	AC129706 Rattus no
42	45	4.5	174248	2	AC113654	AC113654 Rattus no
43	45	4.5	183334	2	AC106174	AC106174 Rattus no
44	44.8	4.5	54677	2	AC108286	AC108286 Rattus no
45	44.8	4.5	55365	2	AC095743	AC095743 Rattus no

ALIGNMENTS

RESULT 1  
HSA297538  
LOCUS Homo sapiens partial RARA gene, intron 2. linear PRI 06-JAN-2001  
DEFINITION  
ACCESSION AJ297538  
VERSION AJ297538.1 GI:12054223  
KEYWORDS Rara gene; retinoic acid receptor alpha.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 16913)  
AUTHORS Cross,N.C.P. and Reiter,A.  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 16913)

**AUTHORS** Cross, N. C. P.  
**TITLE** Direct Submission  
**JOURNAL** Submitted (07-AUG-2000) Cross N. C. P., Department of Haematology, Imperial College School of Medicine, Hammersmith Hospital, London,

**FEATURES** Location/Qualifiers

```

source
1. .16913
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"

```

gene  
1. .16913  
/map-17q12

intron  
1. .16913  
/yelles- RNA

```
/genre=RARA-
/number=2
```

```
repeat_region 1036. .194/  
/rpt_family=
```

```
repeat_region 1949. .2233
/rpt_family=
```

```
repeat_region 2236. 2536
/rpt_family=
```

```
repeat_region 4008. .4315
/rpt_family=
```

```
repeat_region 4327. .4619
/rpt family=
```

```
repeat_region 4669. .4975
/rpt family=
```

```
repeat_region 4982..5009
               /note="TTTA"
```

```
repeat_region 5011.528/
/rpt family=
```

```
repeat_region 528/.559/  
/rpt_family=
```

```
repeat_region 5666..5776
/rpt family=
```

```
repeat_region /090.: /148
/rpt_family=
```

```
repeat_region /rpt_family=
```

```
repeat_region /331.:7041
/rpt_family=
```

```
repeat_region 0516:.0005  
/rpt_family=
```

```
repeat_region      8833.79194  
/rpt_family=
```

```
repeat_region
10143: .1019
/note="(CGGG
```

IN  
COUNTY OF 4/30 C

Every Match 100.08%

atches 1001; Conservative

1 TGATCTCCCTCCCCCACCAGT

15105 TGATCTCCCCCTCCCCCACCAGT

61 AAAGTTAGCACAAACAAGGGCTG

15165 AAAGTTAGCACACAAGGGCTG

121 AGGCCCCCAACCCGCCCCCCC

15225 AGGCCCCCAACCCGCCCCCCC

181 CTTGTCTGGCCTGGGGAGAAATGAG

15285 CTTGTCTGGCCTGGGAGAAATGAC

241 GTGAAGAGCAGACGGCGGTGGGG

\_\_\_\_\_

Query Match	100.0%;	Score 1001;	DB 9;	Length 16913;
Best Local Similarity	100.0%;	Ref. No. 3.6e-367;		
Matches 1001;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY 1	TGATCTCCCTCC	CCCCCACCACG	CTCGATCTCTATTG	TCTTACTGCTTTTACGTCCTTGCAA 60
DB				
15105	TGATCTCCCTCC	CCCCCACCACG	CTCGATCTCTATTG	TCTTACTGCTTTTACGTCCTTGCAA 15164
QY 61	AAAGTTAGCA	CAACAAAGGG	TGCTTTTG	TGCTCTACCCCTCTGCTCTCGCTCGCTCACCC 120
DB				
15165	AAAGTTAGCA	CAACAAAGGG	TGCTTTTG	TGCTCTACCCCTCTGCTCTCGCTCACCC 15224
QY 121	AGGCCCCCA	ACCCGCCGCC	CCAGCAGCTG	TTCTCAGGCCTCTCAGGCCTGCTGATTTG 180
DB				
15225	AGGCCCCCA	ACCCGCCGCC	CCAGCAGCTG	TTCTCAGGCCTCTCAGGCCTGCTGATTTG 15284
QY 181	CTTCTCTGG	CTGGGAGA	ATGAGTGGG	AGAAACACAGGCCAGGCGAGTTGGTGTGGA 240
DB				
15285	CTTCTCTGG	CTGGGAGA	ATGAGTGGG	AGAAACACAGGCCAGGCGAGTTGGTGTGGA 15344
QY 241	GTGAAGAGC	ACAGCGCG	TGGGAGGTC	GAGAGAGAATCTGCTGGGCTGGGGATGGTGTG 300

Db	15345	GTGAAGAGCAGACGCGGTGGGGAGGTCAGGAGAGAAATCTGCTGGGCTGGGGATGGTG	15404
Qy	301	GGCATCAACTGTCCTATTTGCTGCAGGCTGGTCTTGGGCGAGGAAGGAGTGGGGGGCCA	360
Db	15405	GGCATCAACTGTCCTCATTTGCTGCAGGCTGGTCTTGGGCGAGGAAGGAGTGGGGGGCCA	15464
Qy	361	TAGCAGTCTGGTTCAGCCAGGCTGBCCTGGGAAGTGGTGCCAGGCACCTACTAAGAGCCA	420
Db	15465	TAGCAGTCTGGTTCAGCCAGGCTGBCCTGGGAAGTGGTGCCAGGCACCTACTAAGAGCCA	15524
Qy	421	GGAAAGCCCTGCCAAGGTTGTTGGGCTAGTTCCTCTGTCATCAGCGCCTCAGCAGCCCCA	480
Db	15525	GGAAAGCCCTGCCAAGGTTGTTGGGCTAGTTCCTCTGTCATCAGCGCCTCAGCAGCCCCA	15584
Qy	481	CTGCTCTGTCAGGTAAGGGGGAGGGTGTAGCACATAGTCAGCCCTCTGGTGTCCCATG	540
Db	15585	CTGCTCTGTCAGGTAAGGGGGAGGGTGTAGCACATAGTCAGCCCTCTGGTGTCCCATG	15644
Qy	541	CTTCCTCTCTGTGCCCAATTTTAGGGCCATGTGATTTGGGGCTATGTGACTCATGTC	600
Db	15645	CTTCCTCTCTGTGCCCAATTTTAGGGCCATGTGATTTGGGGCTATGTGACTCATGTC	15704
Qy	601	TGTAAGTGCTTTGGGCCAGGAGCTGTGGCACCTTTAAATGCCAGCAGTCTCATGTGCC	660
Db	15705	TGTAAGTGCTTTGGGCCAGGAGCTGTGGCACCTTTAAATGCCAGCAGTCTCATGTGCC	15764
Qy	661	GGAGTTTGGGGTAGGGCTAGGTAGTGTGGAATATGGGAGGAGCAGGATCTGTCTA	720
Db	15765	GGAGTTTGGGGTAGGGCTAGGTAGTGTGGAATATGGGAGGAGCAGGATCTGTCTA	15824
Qy	721	CCTAGGAGGATCTCATCTCATCTTGCCCTTGACAAAGAACTTCAACCTTGGTAGG	780
Db	15825	CCTAGGAGGATCTCATCTCATCTTGCCCTTGACAAAGAACTTCAACCTTGGTAGG	15884
Qy	781	GGCCTCAGGACATGCTCGTGGCCCCCTTGGGAATCTGGGATGTCTGGTCAATATTTCT	840
Db	15885	GGCCTCAGGACATGCTCGTGGCCCCCTTGGGAATCTGGGATGTCTGGTCAATATTTCT	15944
Qy	841	TATCTGCACCCACACCCCTTAGCTGCCAGGCTTTGGACATGGATAGCCCTACCCAAC	900
Db	15945	TATCTGCACCCACACCCCTTAGCTGCCAGGCTTTGGACATGGATAGCCCTACCCAAC	16004
Qy	901	CCAGCCCTGTTCTGCCTACAGTGTAGTGAGCCATGGAGCCACAGACACTGGGAGGATTTGGCCA	960
Db	16005	CCAGCCCTGTTCTGCCTACAGTGTAGTGAGCCATGGAGCCACAGACACTGGGAGGATTTGGCCA	16064
Qy	961	GTGAGGGCTGCCCTGCTGCTGTGGGTCAACCCCTCTGGCTG	1001
Db	16065	GTGAGGGCTGCCCTGCTGCTGTGGGTCAACCCCTCTGGCTG	16105

RESULT 2	
AC090426/c	
LOCUS	138999 bp DNA linear PRI 29-MAY-2002
DEFINITION	Homo sapiens chromosome 17 clone 205m17 map 17q21.1, complete sequence.
ACCESSION	AC090426
VERSION	AC090426.1 GI:12965315
KEYWORDS	HTG.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 138999) Gu, B., Xiong, H., Zhou, Y., Chen, B., Lu, L., Zhong, M., Yin, H., Huang, W., Ren, S., Chen, Z., and Fu, G.
TITLE	Variant-type PML-RAR(alpha) fusion transcript in acute promyelocytic leukemia: use of a cryptic coding sequence from intron 2 of the RAR(alpha) gene and identification of a new clinical subtype resistant to retinoic acid therapy
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 99 (11), 7640-7645 (2002)
MEDLINE	22028997
PUBMED	12032336



```

REFERENCE      2 (bases 1 to 138999)
AUTHORS        Gu.B., Xiong.H., Zhou.Y., Chen.B., Lu.L., Zhong.M., Yin.H.,
                Huang.W., Ren.S., Chen.S.F., Chen.Z. and Fu.G.
TITLE          Direct Submission
JOURNAL        Shanghai, Shanghai, Shanghai 201203, P.R.China
COMMENT        -----Genome Center-----
                Center:Chinese National Human Genome Center at Shanghai
                Center code:CHGC Website: http://www.chgc.sh.cn Contact:
                fugang@chgc.sh.cn.
FEATURES
source
1. .138999
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="17"
   /map="17q21.1"
   /clone="203ml17"
   387. .483
   /rpt_family="SINE/MIR"
   /rpt_unit=MIR
repeat_region  725. .1032
   /rpt_family="SINE/Alu"
   /rpt_unit=AluB
repeat_region  1033. .1198
   /rpt_family="SINE/Alu"
repeat_region  1217. .1514
   /rpt_family="SINE/Alu"
   /rpt_unit=AluSp
repeat_region  complement(1534. .1754)
   /rpt_family="SINE/MIR"
   /rpt_unit=MIR
repeat_region  2291. .2366
   /rpt_family="SINE/MIR"
   /rpt_unit=MIR
repeat_region  complement(2607. .2665)
   /rpt_family="SINE/MIR"
   /rpt_unit=MIR
repeat_region  3651. .3905
   /rpt_family="LINE/L1"
   /rpt_unit=L1ME2
repeat_region  3977. .4029
   /rpt_family="LINE/L1"
   /rpt_unit=L1ME2
repeat_region  4030. .4213
   /rpt_family="LTR/MaLR"
   /rpt_unit=MSTD
repeat_region  4214. .4509
   /rpt_family="SINE/Alu"
   /rpt_unit=AluB
repeat_region  4514. .4811
   /rpt_family="SINE/Alu"
   /rpt_unit=AluSx
repeat_region  4812. .5018
   /rpt_family="LTR/MaLR"
   /rpt_unit=MSTD
repeat_region  5019. .5102
   /rpt_family="LINE/L1"
   /rpt_unit=L1ME2
repeat_region  5236. .5443
   /rpt_family="SINE/MIR"
   /rpt_unit=MIR
repeat_region  5467. .5699
   /rpt_family="SINE/MIR"
   /rpt_unit=MIR
repeat_region  6547. .6708
   /rpt_family="SINE/MIR"
   /rpt_unit=MIR
repeat_region  6827. .6917
   /rpt_family="Low_complexity"
   /rpt_unit=GA-rich
repeat_region  7733. .7779
   /rpt_family="SINE/MIR"
   /rpt_unit=MIR
8268. .8411
   /rpt_family="Simple_repeat"
   /rpt_unit=(TGGG)n
9224. .9532
   /rpt_family="SINE/Alu"
   /rpt_unit=AluY
9546. .9639
   /rpt_family="Low_complexity"
   /rpt_unit=CT-rich
10025. .10426
   /rpt_family="LTR/Retroviral"
   /rpt_unit=MLT2A
10433. .10471
   /rpt_family="Simple_repeat"
   /rpt_unit=(TA)n
10472. .10527
   /rpt_family="LTR/Retroviral"
   /rpt_unit=MLT2A
complement(10675. .10962)
   /rpt_family="SINE/Alu"
   /rpt_unit=AluY
11013. .11321
   /rpt_family="SINE/Alu"
   /rpt_unit=AluY
complement(11840. .12192)
   /rpt_family="LTR/MaLR"
   /rpt_unit=MSTD
12729. .12835
   /rpt_family="SINE/Alu"
   /rpt_unit=AluY
complement(12841. .12891)
   /rpt_family="LINE/L2"
   /rpt_unit=L2
complement(14228. .14548)
   /rpt_family="SINE/Alu"
   /rpt_unit=AluY
complement(14618. .14701)
   /rpt_family="SINE/MIR"
   /rpt_unit=MIR
14953. .15255
   /rpt_family="SINE/Alu"
   /rpt_unit=AluJo
15311. .15473
   /rpt_family="Simple_repeat"
   /rpt_unit=(TA)n
15669. .15981
   /rpt_family="SINE/Alu"
   /rpt_unit=AluSx
16316. .16353
   /rpt_family="Simple_repeat"
   /rpt_unit=(TAAA)n
complement(16863. .16910)
   /rpt_family="SINE/MIR"
   /rpt_unit=MIR
17026. .17324
   /rpt_family="SINE/Alu"
   /rpt_unit=AluSg
17337. .17618
   /rpt_family="SINE/Alu"
   /rpt_unit=AluSx
17619. .18024
   /rpt_family="LTR"
   /rpt_unit=LTR54
complement(18026. .18086)
   /rpt_family="SINE/MIR"
   /rpt_unit=MIR
complement(18292. .18607)
   /rpt_family="SINE/Alu"
   /rpt_unit=AluYb8
complement(18948. .19264)
   /rpt_family="SINE/Alu"
   /rpt_unit=AluSx
complement(20012. .20168)

```



















QY 721 CCTAGGAGGACCTCCTCATCCATCCCTTGGCCCTGGCAACAGACACTTGAACCTTGGTAGG 780  
Db 42395 CCTAGGAGGACCTCCTCATCCATCCCTTGGCCCTGGCAACAGACACTTGAACCTTGGTAGG 42454  
QY 781 GGCTCAGGACGATGCTGCTGGCCCTTGGGAATCTGGGATGCTCCTGCTCATAGTTCT 840  
Db 42455 GGCTCAGGACGATGCTGCTGGCCCTTGGGAATCTGGGATGCTCCTGCTCATAGTTCT 42514  
QY 841 TATCTTGCACCCCAACACCTTACCTGCCCCAGGCTTTGGACATGGATGCCCTTACCCCAAC 900  
Db 42515 TATCTTGCACCCCAACACCTTACCTGCCCCAGGCTTTGGACATGGATGCCCTTACCCCAAC 42574  
QY 901 CCAGCCCTGCTCTGCTACAGTATGGCATGGGATGGAGGATGGGAGATTTGGCCA 960  
Db 42575 CCAGCCCTGCTCTGCTACAGTATGGCATGGGATGGAGGATGGGAGATTTGGCCA 42634  
QY 961 GTGAGGCTGCCCTGCTGCTGGGTCACCCCTCTGGCTG 1001  
Db 42635 GTGAGGCTGCCCTGCTGCTGGGTCACCCCTCTGGCTG 42675  
RESULT 8  
AX344790 13508 bp DNA linear PAT 01-FEB-2002  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
Location/Qualifiers  
1. .13508  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 2350 a 286 c 4030 g 6841 t 1 others  
ORIGIN

Query Match 58.4%; Score 584.2; DB 6; Length 13508;  
Best Local Similarity 74.7%; Pred. No. 2.5e-151;  
Matches 733; Conservative 0; Mismatches 248; Indels 0; Gaps 0;  
QY 21 AGTCTGATGCTATGTTACTGCTTTTACGCTCTGGAAAAGTTAGCACACAAAGGG 80  
Db 11719 AGTCTGATGCTATGTTACTGCTTTTACGCTCTGGAAAAGTTAGTATAAAGGG 11778  
QY 81 CTGCTTTGGCTCACCCCTCTGCTCTGCTCCGCTCACCCAGCCGCCCAACCCGCCCC 140  
Db 11779 TTGTTTGGGTTATTTTGGTTTGGTTTATTTAGGTTTATTTAGTTTATTTTCGTTT 11838  
QY 141 CCAGACGCTGTTCTCAGCCCTCTACGCTGCTGATTTGCTGCTGGCTGGGAGAA 200  
Db 11839 TTTAGTAGTTGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTTTATGTT 11898  
QY 201 TGAGTGGGAGAAACACGCGCAGGCGAGTGGTGGTGGAGTGAAGACAGACGCGGTG 260  
Db 11899 TGAGTGGGAGAAATAGGTTAGGTTAGTGGTGGTGGAGTGAAGACAGACGCGGTG 11958  
QY 261 GGGAGTCCAGAGAGAAATAGTGGCTGGGATGGTGGGATGCTGCTGCTGCTGCTGCT 320  
Db 11959 GGGAGTCCAGAGAGAAATAGTGGCTGGGATGGTGGGATGCTGCTGCTGCTGCTGCT 12018  
QY 321 TGAGGCTGCTCTTGGGAGGAGGATGGGAGGATGGGAGGATGGGAGGATGGGAGGAT 380  
Db 12019 TGAGGCTGCTCTTGGGAGGAGGATGGGAGGATGGGAGGATGGGAGGATGGGAGGAT 12078

QY 381 GCTGCCCTGGGAAGTGGTCCAGGCACTACTAAGAGCCAGAAAGCCCTGCCAAGGTTG 440  
Db 12079 GTTGGTTGGGAAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 12138  
QY 441 TTGGCTTAGTTCCCTCTCATCAGCGCCCTAGCAGCCCACTGCTGTCAGAGTAAAGGG 500  
Db 12139 TTGGCTTAGTTCCCTCTCATCAGCGCCCTAGCAGCCCACTGCTGTCAGAGTAAAGGG 12198  
QY 501 GGAGGGTGGTAGACATAGTACAGCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 560  
Db 12199 GGAGGGTGGTAGACATAGTACAGCCCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 12258  
QY 561 ATTTTAGGCGCATGATGCTGGGCTAGTACATCATCTGCTGTAAGTGGTGGTGGTGGT 620  
Db 12259 ATTTTAGGCGCATGATGCTGGGCTAGTACATCATCTGCTGTAAGTGGTGGTGGTGGT 12318  
QY 621 AGCTGTGGGCACTTTAAATGCCAGCCCTCTCATGCTGCCGAGTGGTGGGCTAGGCTAG 680  
Db 12319 AGCTGTGGGCACTTTAAATGCCAGCCCTCTCATGCTGCCGAGTGGTGGGCTAGGCTAG 12378  
QY 681 GTAGGATTTGGGAATAGGGAGGAGGATCTGCTACCTAGGAGGATCTCTATC 740  
Db 12379 GTAGGATTTGGGAATAGGGAGGAGGATCTGCTACCTAGGAGGATCTCTATC 12438  
QY 741 CATCTTGGCCCTGGACAAGAGAACTTCAACGTTGGTAGGGCCCTCAGGACGATGCTCG 800  
Db 12439 TATTTTGGGTTTGGATAAGAGAAATTTGAACGTTGGTAGGGCCCTCAGGACGATGCTCG 12498  
QY 801 TGGCCCTTGGGAATCTGGGATTTGCCCTGCTCATAGTTCTTATCTTGCACCAACACCT 860  
Db 12499 TGGCTTTTGGGAATTTGGGATTTGGGATTTGGTATAGTTATTTATTTATTTATTTAT 12558  
QY 861 TAGCTGCCAGGCTTTGGACATGGATGCCCTTACCAACCCAGCCCTGTTCTGCCTACA 920  
Db 12559 TAGTTTGGTATTTGGATATGGATGTTTATTTATTTATTTATTTATTTATTTATTTAT 12618  
QY 921 GTGATGGGATGGAGGACACTGGGAGGATTTGGGAGGATTTGGGAGGATTTGGTCTGT 980  
Db 12619 GTGATGGGATGGAGGATTTGGGAGGATTTGGGAGGATTTGGTCTGT 12678  
QY 981 CTGGGTCACCCCTCTCTGGCTG 1001  
Db 12679 TTGGGTTATTTTGGTTG 12699

RESULT 9  
AX344791/c 13508 bp DNA linear PAT 01-FEB-2002  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
FEATURES  
source  
Location/Qualifiers  
1. .13508  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 3338 a 286 c 3789 g 6094 t 1 others  
ORIGIN  
Query Match 50.7%; Score 507.2; DB 6; Length 13508;  
Best Local Similarity 69.2%; Pred. No. 6.4e-130;  
Matches 692; Conservative 0; Mismatches 308; Indels 0; Gaps 0;  
QY 1 TGATCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 60

```
Db 1810 TAATCTCCCTCCGCCCACTAAATATATCTATTATCTACTTTTACGTCCTTAAAA 1751
Qy 61 AAAGTTAGCACACAAAGGCTGCTTTGGGCTCACCCTCTGCTCTGCTCGCTCACC 120
Db 1750 AAAATTAACACAAACAAACAACTACTTTTAACTACCCCTCTACTCTCAACCTCACC 1691
Qy 121 AGGCCCCCAACCCCGCCGCCAGCAGCTGTTCTCAGGCGCTCTCAGCTGCTGATTGG 180
Db 1690 AAACCCCAACCCCGCCGCCCACTACTTCTCAACCTCTCAACCTACTTAATTTA 1631
Qy 181 CTTGCTCGCTGGGAGAGTGAAGTGGGAGAAACACAGGCGAGGCTGTTGGTGTGA 240
Db 1630 CTTATCTACCTTAAACAAATATAAATAAATAAATAAATAAATAAATAAATAAATAA 1571
Qy 241 GTGAGACAGACGCGCTGGGAGGTGAGGAGAGTCTGAGGAGAACTGCTGGGCTGGGATGGTGTG 300
Db 1570 ATAAAAAACAACGACGATAAAAAAATCAAAAAAATACTTAACTAAAAAATATATA 1511
Qy 301 GGCATCAACTGTCCATTTGCTGAGGCTGCTTGGGCGAGGAAAGGAGGAGGAGGAGG 360
Db 1510 AACATCACTATCCATTTACTACAACCTAACTTAACTTAAACAAAAAATAAATAAATAA 1451
Qy 361 TAGCAGTGTGCTGACGAGGCTGCTGCGGAGTGGTGGAGTGGTGGGAGGCTACTTAAGGCCA 420
Db 1450 TAACAATACTAATCAACCAAACTAACTTAAATAAATAAATAAATAAATAAATAAATAA 1391
Qy 421 GGAAGCCCTGCCAGGTTGTTGGCTAGTTCCTGTCATGAGCGCTAGAGGCCCA 480
Db 1390 AAAAACCCCTACCAAAATTTATTAACCTAATTCCTTATCAACGCGCTTAACAAACCCCA 1331
Qy 481 CTGCTCTGCTGAGTGAAGGGGAGGCTGTGACACATAGTACGCGCTGCTGCTGCCAGT 540
Db 1330 CTATATCTACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1271
Qy 541 CTTCTCTCTCTGTGCCCCCAATTTTAGGCGCATGTGATTTGGGCTATGTGACTCATGTC 600
Db 1270 CTTCTCTCTCTATACCCCAATTTTAAACCACTAATAATTTAAACATATATACTCATATC 1211
Qy 601 TGTAGGTGCTTGGCCAGGAGCTGTGGCCACCTTTAATGCGCCAGCTGCTCATGTGCC 560
Db 1210 TATAAATACTTAAACCAAACTAATAACACCTTTAATAATCAACCACTCATATATACC 1151
Qy 651 GGAGTTTGGGCTAGGCTAGGTAGGATTTGGAATATGGGAGGAGGAGGATCTGTCTA 720
Db 1150 GAATTTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 1091
Qy 721 CCTAGGAGGATCCTCATCCATCCTTGGCCCTGGACAGGAGAACTTGAACCTTGGTAGG 780
Db 1090 CCTAAAAAACATCCTCATCCATCCTTAACCTTAAACCAAAACAACTTAAACCTTAAATAA 1031
Qy 781 GGCCTCAGGAGATGCTCGTGGGCCCTTGGGAATCTGGGATTTGCTGGTGCATAGTTCT 840
Db 1030 AACCTCAAAAGATACCTACGTAACCCCTTAAATAATCTTAAATAATCTTAAATAATCT 971
Qy 841 TATCTTGACCAACACACCTTAGCTGCCAGGCTTTGGACATGGATAGCCCTACCCAC 900
Db 970 TATCTTACACCAACACCTTAACTACCCAACTTTAAACATAAATAAATAAATAAATAA 911
Qy 901 CCAGCCCTGTTCTGCTACAGTGTGGCATGGAGCCAGACACTGGGAGGATTTGGCCA 960
Db 910 CCAACCCCTATTCTACCTACAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 851
Qy 961 GTAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1000
Db 850 ATAAAAACTACCCCTACTATCTAATAATACCCCTCCTTAAC 811
```

RESULT 10  
HUMUT5218  
LOCUS  
DEFINITION  
ACCESSION

HUMUT5218 516 bp DNA linear STS 26-SEP-1995  
Human STS UT5218, sequence tagged site.  
L17737

VERSION L17737.1 GI:957208  
KEYWORDS STS; STS sequence; primer; sequence tagged site.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 516)  
AUTHORS Gerken,S.C., Matsunami,N., Lawrence,E., Carlson,M., Moore,M.,  
Ballard,L., Melis,R., Robertson,M., Bradley,P., Elsner,T.,  
Tinsley,A., Rodriguez,P., Albertsen,H., Lalouel,J.-M. and White,R.  
Genetic and physical mapping of simple sequence repeat containing  
sequence tagged sites from the human genome  
Unpublished (1993)  
COMMENT On Aug 25, 1995 this sequence version replaced gi:308251.  
Submitted by: Utah Center for Human Genome Research University of  
Utah, Dept. of Human Genetics  
2160 Eccles Institute of Human Genetics  
Salt Lake City, UT 84112  
e-mail: sts@corona.med.utah.edu  
Primer A: TGCTGACGCTGCTCTGG  
Primer B: GCGGCTGATGACAGGAA  
32P-label: Primer  
PCR Profile:  
Initial Denaturation: 94C 300sec  
PCR Cycles: 30  
Denaturation: 94C 10sec  
Annealing: 62C 10sec  
Extension: 72C 20sec  
Mg++: 1.5mM  
Gel: Acrylamide 7%, Formamide 32%, Urea 34%  
Alleles: 1.  
FEATURES  
Location/Qualifiers  
1..516  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
171..321  
/standard\_name="STS UT5218"  
primer\_bind 171..189  
primer\_bind complement(303..321)  
BASE COUNT 95 a 105 c 163 g 128 t 25 others  
ORIGIN

Query Match 38.4%; Score 384.2; DB 11; Length 516;  
Best Local Similarity 89.1%; Pred. No. 1.1e-95;  
Matches 457; Conservative 0; Mismatches 47; Indels 9; Gaps 5;

Qy 149 CTGTTCTCAGGCTCTCAGCCTCTCTG-ATTGCTTGTCTGGCTGGGAGATGAGGTG 207  
Db 1 CTGCTCTNAGGCTCTNAGCCTGTTGAATTTGCTTGTGGCTGGGAGATGAGGTG 60  
Qy 208 GGAGAAAACAGGCCAGGCGAGTGTGGTGTGAGTGAAGAGCAGACGCGGTGGGAGGT 267  
Db 61 GGAGAAAACAGGCCAGGCGAGTGTGGTGTGAGTGAAGAGCAGACGCGGTGGGAGGT 120  
Qy 268 CAGGAGAGATCTGCTGGCTGGGATGCTGGGATCACTGTCCATTGCTGCAGGC 327  
Db 121 CAGGAGAGATCTNCTGGCTGGGATGCTGGGATCACTGTCCATTGCTGCAGGC 180  
Qy 328 TGGTCTTGGGCGAGGAAGGGATGGGGCCATAGACGTGCTGGTCAAGCAGCTGGCC 387  
Db 181 TGGTCTTGGGCGAGGAAGGGATGGGGCCATAGACGTGCTGGTCAAGCAGCTGGCC 240  
Qy 388 TGGGAGTGTGCTCCAGGCACCTACTAAGAGCCAGGAAGCCCTGCCAAGCTGTGTCCT 447  
Db 241 TNGGAAGTGTGCTCCAGGCACCTACTAAGAGCCAGGAAGCCCTGCCAAGCTGTGTCCT 300  
Qy 448 AGTTCCCTGTCTATCAGCCCTCTAGCAGCCCCCTACTGTCTGCAGGTAAGGGGGAGG- 506  
Db 301 AGTTCCCTGTCTATCAGCCCTCTAGCAGCCCCCTACTGTCTGCAGGTAAGGGGGAGG 360  
Qy 507 TGGTAGCACATAGTCAGCCCTCT--GGTGTCCCATCTCTCTCTCTGTGCCCAATTT 564  
Db 361 TGGTAGCACATAGTCAGCCCTCTGGGTGTTNCCATGTTCTTCTCTCTGTGCCCAATTT 420





ORGANISM      Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE      1 (bases 1 to 914)



of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 194929)  
 Worley, K.C.  
 Direct Submission  
 Submitted (18-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 14, 2002 this sequence version replaced gi:20303396.

REFERENCE  
 AUTHORS  
 TITLE  
 JOURNAL

COMMENT

----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GTMP  
 Center clone name: CH230-112G1  
 ----- Summary Statistics  
 Sequencing vector: Plasmid;  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 128159 bases at least Q40  
 Consensus quality: 135188 bases at least Q30  
 Consensus quality: 139965 bases at least Q20  
 -----

\* NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 71 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1036: contig of 1036 bp in length  
 1037 1136: gap of unknown length  
 1137 2399: contig of 1263 bp in length  
 2400 2499: gap of unknown length  
 2500 3523: contig of 1024 bp in length  
 3524 3623: gap of unknown length  
 3624 5005: contig of 1382 bp in length  
 5006 5105: gap of unknown length  
 5106 6355: contig of 1250 bp in length  
 6356 6456: gap of unknown length  
 6457 7907: contig of 1452 bp in length  
 7908 8008: gap of unknown length  
 8009 9317: contig of 1310 bp in length  
 9318 9417: gap of unknown length  
 9418 10644: contig of 1227 bp in length  
 10645 10744: gap of unknown length  
 10745 11925: contig of 1181 bp in length  
 11926 12025: gap of unknown length  
 12026 13524: contig of 1499 bp in length  
 13525 13625: gap of unknown length  
 13626 14677: contig of 1053 bp in length  
 14678 14777: gap of unknown length  
 14778 15855: contig of 1078 bp in length  
 15856 15955: gap of unknown length  
 15956 17320: contig of 1365 bp in length  
 17321 17421: gap of unknown length  
 17422 18455: contig of 1035 bp in length  
 18456 18555: gap of unknown length  
 18556 20137: contig of 1582 bp in length  
 20138 20237: gap of unknown length  
 20238 21836: contig of 1599 bp in length  
 21837 21936: gap of unknown length  
 21937 23431: contig of 1495 bp in length  
 23432 23531: gap of unknown length  
 23532 24970: contig of 1439 bp in length  
 24971 25070: gap of unknown length  
 25071 26350: contig of 1279 bp in length  
 26351 26449: gap of unknown length

26450 27905: contig of 1456 bp in length  
 27906 28005: gap of unknown length  
 28006 29330: contig of 1325 bp in length  
 29331 29430: gap of unknown length  
 29431 30576: contig of 1146 bp in length  
 30577 30676: gap of unknown length  
 30677 32160: contig of 1484 bp in length  
 32161 32260: gap of unknown length  
 32261 33426: contig of 1166 bp in length  
 33427 33526: gap of unknown length  
 33527 35660: contig of 2134 bp in length  
 35661 35760: gap of unknown length  
 35761 37766: contig of 2006 bp in length  
 37767 37866: gap of unknown length  
 37867 39093: contig of 1227 bp in length  
 39094 39193: gap of unknown length  
 39194 41094: contig of 1901 bp in length  
 41095 41194: gap of unknown length  
 41195 42396: contig of 1202 bp in length  
 42397 42496: gap of unknown length  
 42497 43637: contig of 1141 bp in length  
 43638 43737: gap of unknown length  
 43738 45693: contig of 1956 bp in length  
 45694 45793: gap of unknown length  
 45794 47577: contig of 1784 bp in length  
 47578 47677: gap of unknown length  
 47678 49368: contig of 1691 bp in length  
 49369 49468: gap of unknown length  
 49469 50928: contig of 1460 bp in length  
 50929 51028: gap of unknown length  
 51029 53617: contig of 2589 bp in length  
 53618 53718: gap of unknown length  
 53719 55954: contig of 2237 bp in length  
 55955 56054: gap of unknown length  
 56055 58165: contig of 2111 bp in length  
 58166 58265: gap of unknown length  
 58266 60797: contig of 2531 bp in length  
 60798 60897: gap of unknown length  
 60898 63491: contig of 2595 bp in length  
 63492 63591: gap of unknown length  
 63592 65691: contig of 2100 bp in length  
 65692 65791: gap of unknown length  
 65792 67770: contig of 1979 bp in length  
 67771 67870: gap of unknown length  
 67871 70084: contig of 2214 bp in length  
 70085 70184: gap of unknown length  
 70185 72274: contig of 2090 bp in length  
 72275 72374: gap of unknown length  
 72375 74673: contig of 2299 bp in length  
 74674 74773: gap of unknown length  
 74774 77271: contig of 2498 bp in length  
 77272 77371: gap of unknown length  
 77372 79490: contig of 2119 bp in length  
 79491 79590: gap of unknown length  
 79591 81823: contig of 2233 bp in length  
 81824 81923: gap of unknown length  
 81924 83627: contig of 1704 bp in length  
 83628 83727: gap of unknown length  
 83728 85740: contig of 2013 bp in length  
 85741 88540: gap of unknown length  
 88541 88656: contig of 2816 bp in length  
 88657 88756: gap of unknown length  
 88757 91337: contig of 2581 bp in length  
 91338 91437: gap of unknown length  
 91438 93224: contig of 2487 bp in length  
 93225 94024: gap of unknown length  
 94025 96898: contig of 2874 bp in length  
 96899 99470: contig of 2472 bp in length

Query Match 5.5%; Score 55; DB 2; Length 194929;  
 Best Local Similarity 45.8%; Pred. No. 0.00028;  
 Matches 162; Conservative 0; Mismatches 191; Indels 1; Gaps 1;





GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 07:57:38 : Search time 196.299 seconds  
(without alignments)  
11483.757 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_5000\_6000  
Perfect score: 1001  
Sequence: 1 tgatccctcccccacc.....tgggtaccctctggtgctg 1001

Scoring table: IDENTITY\_NUC  
Gap 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues  
Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :			
N_Geneseq_101002.*			
1:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*		
2:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*		
3:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*		
4:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*		
5:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*		
6:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*		
7:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*		
8:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*		
9:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*		
10:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*		
11:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*		
12:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*		
13:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*		
14:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*		
15:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*		
16:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*		
17:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*		
18:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*		
19:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*		
20:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*		
21:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*		
22:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*		
23:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*		
24:	/SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1001	100.0	20512	24	Genomic DNA encodi
2	584.2	58.4	13508	24	Human 'chemically m
3	507.2	50.7	13508	24	Human chemically m
4	44.6	4.5	549	13	Sequence encoding
5	44.6	4.5	549	13	AAQ20379
6	42.6	4.3	357	22	Human polynucleoti
7	42.6	4.3	34980	22	C glutamicum codin
8	42.2	4.2	30032	22	Human nervous syst
9	42	4.2	37664	22	Human immune/haema

C 10	42	4.2	61710	22	AAK83782
C 11	41.4	4.1	9359	24	ABK10687
C 12	41.4	4.1	9359	24	ABK11039
C 13	39.6	4.0	1894	13	AAQ20466
C 14	39.6	4.0	1894	18	AAAT71220
C 15	39.2	3.9	5786	22	AAS29976
C 16	38.4	3.8	1407	21	AAZ89275
C 17	38.4	3.8	2237	22	AAS30201
C 18	38.4	3.8	2378	21	AAAD01016
C 19	38.4	3.8	2620	21	AAAS3824
C 20	38.4	3.8	3879	21	AAZ89276
C 21	38.2	3.8	6865	22	AAAF80287
C 22	38.2	3.8	10003	22	AAAF80295
C 23	38.2	3.8	114793	22	AAAD08215
C 24	38	3.8	1337	20	AAZ17263
C 25	38	3.8	1501	21	AAAC87194
C 26	38	3.8	1501	21	AAAC89325
C 27	38	3.8	4032	19	AAV23237
C 28	37.6	3.8	410	22	AAI82055
C 29	37.6	3.8	600	24	ABQ52496
C 30	37.6	3.8	600	24	ABQ52497
C 31	37.6	3.8	10732	21	AAAI0594
C 32	37.4	3.7	5121	22	AAK83937
C 33	37.4	3.7	5864	18	AAAT72812
C 34	37.4	3.7	53522	24	AAAD30228
C 35	37.4	3.7	53526	19	AAAT94101
C 36	37.4	3.7	53577	17	AAAT18551
C 37	37.4	3.7	53577	19	AAAT94108
C 38	37.2	3.7	840	24	ABQ35494
C 39	37.2	3.7	993	24	ABK35988
C 40	37.2	3.7	2042	24	ABN95855
C 41	37.2	3.7	801	22	AAH92234
C 42	37	3.7	1521	22	AAAD08644
C 43	37	3.7	1578	23	AAAS85330
C 44	37	3.7	2380	22	AAAS14817
C 45	37	3.7			

ALIGNMENTS

RESULT 1	
AAL38339	
ID	AAL38339 standard; DNA; 20512 BP.
XX	AAL38339;
AC	
XX	
DT	15-AUG-2002 (first entry)
DE	Genomic DNA encoding the human nuclear hormone receptor protein.
XX	
KW	Human nuclear hormone receptor; therapeutic agent; gene therapy;
KW	Immune response; chromosome 17; single nucleotide polymorphism;
KW	gene; ds.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	1684..15208
FT	/tag= a
FT	/product= "Human nuclear hormone receptor protein"
FT	1684..1846
FT	/tag= b
FT	/number= 1
FT	1847..6808
FT	/tag= c
FT	/number= 1
FT	6809..6957
FT	/tag= d
FT	/number= 2
FT	6958..8277
FT	/tag= e
FT	/number= 2







Db 910 CCAACCCCTATTCTACCTACATAATAAACAATAAAACCAACACATAAAAAATTTAACA 851  
QY 961 GTGAGGCTGCCCTGCTGCTGGTCCACCCCTCTCTGCT 1000  
Db 850 ATAAAACTACCCCTACTATCTAAATCACCCCTCTCTACT 811

RESULT 4  
AAO20379/c  
ID AAO20379 standard; cDNA; 549 BP.  
XX AC AAO20379;  
XX DT 13-APR-1992 (first entry)  
XX DE Sequence encoding porcine C-type natriuretic peptide (CNP) contg.  
XX DE precursor.  
XX KW Natriuretic; hypotensive; hormone; ss.  
XX OS Pig.  
XX FH Key Location/Qualifiers  
XX FT CDS 363..524  
XX FT /\*tag= a  
XX PN EP466174-A.  
XX PD 15-JAN-1992.  
XX PF 12-JUL-1991; 91EP-0111629.  
XX PR 13-JUL-1990; 90JP-0186582.  
XX PA (MATS/) MATSUO H.  
XX PI Matsuo H, Kangawa K, Minamino N;  
XX PI WPI; 1992-017805/03.  
XX DR P-PSDB; AAR20182.  
XX PT New C-type natriuretic peptide(s) isolated from porcine brain -  
XX PT possess natriuretic and hypotensive activity  
XX PS Disclosure; Fig 6; 20pp; English.  
XX CC CNP can be isolated from porcine brains using antibodies to CNP-22  
XX CC (JA105047/90) and sequenced. The 22 and 53 C-terminal AAs have  
XX CC natriuretic and hypotensive activities, and are claimed (see AAR20182).  
XX SQ Sequence 549 BP; 85 A; 212 C; 178 G; 74 T; 0 other;

Query Match 4.5%; Score 44.6; DB 13; Length 549;  
Best Local Similarity 49.5%; Pred. No. 0.021;  
Matches 145; Conservative 0; Mismatches 144; Indels 4; Gaps 1;

QY 89 TGCTCAGGAGAAACCCAGGCGAGTGTGTTGGAGTGAAGAGACAGCGCGGTGGGA 148  
Db 306 TGTGCGCCCTTCTTGACCGCGCGAGCTGGGCTCGGCCACCTCTCCCTGGCG 247  
QY 149 CTGTTCTCAGGCTCTCAGC----CTGCTGATTTGCTTCTGCTGGGAGAAATGAG 204  
Db 246 GAGTTTCAGAGGAGCTTCCGGCGCTCCGGGCTTGGCTTCGAGGGCGGAGAGGA 187  
QY 205 GTGGGAGAAACCCAGGCGAGTGTGTTGGAGTGAAGAGACAGCGCGGTGGGA 264  
Db 186 GCGTGAAG 127  
QY 265 GGTGAGAGAGAAATCTGCTGGGCTGGGAGTGGTGTGGGATCAACTGTCCCATTTGCTGCA 324  
Db 126 GGGCGCAGATCGCGGAGAGAGTGTGACGGGGCGGCGGAGAGAGAGAGAGAGAG 67  
QY 325 GGCTGTCTTGGGCGAGGAGAGGAGTGGGGGCCATAGCAGTGTGCTGCTCAGC 377

Db 56 CGGAGCAGACTGGCGGCTGGGTCGGGCGGAGTCCAGTCTCGCGGC 14  
RESULT 5  
AAO20467/c  
ID AAO20467 standard; cDNA; 549 BP.  
XX AC AAO20467;  
XX DT 14-APR-1992 (first entry)  
XX DE Sequence encoding porcine C-type natriuretic peptide (CNP) precursor.  
XX KW Hypotensive agent; natriuretic; brain-specific promoter; ss.  
XX OS Sus scrofa domestica.  
XX FH Key Location/Qualifiers  
XX FT exon 144..521  
XX FT /\*tag= a  
XX PN EP466175-A.  
XX PD 15-JAN-1992.  
XX PF 12-JUL-1991; 91EP-0111630.  
XX PR 13-JUL-1990; 90JP-0186583.  
XX PA (SUNR) SUNTORY LTD.  
XX PI Matsuo H, Kangawa K, Minamino N, Tanaka S, Fuchimura K;  
XX PI Tawaragi Y;  
XX DR WPI; 1992-017806/03.  
XX DR P-PSDB; AAR20074.  
XX PT DNA encoding porcine C-natriuretic peptide - or precursor and  
XX PT fragments with natriuretic and hypotensive activities  
XX PS Claim 12; Fig 5; 23pp; English.  
XX CC Genomic and cDNA encoding porcine CNP are claimed. Also claimed are  
XX CC fragments encoding the last 22 AAs (CNP-22) and the last 53 AAs  
XX CC (CNP-53). Proteins encoded are also claimed. The full length DNA  
XX CC includes a promoter which may act in a brain-specific manner so  
XX CC could be used to control expression of other genes in the brain  
XX CC (see AAO20465).  
XX SQ Sequence 549 BP; 85 A; 212 C; 178 G; 74 T; 0 other;

Query Match 4.5%; Score 44.6; DB 13; Length 549;  
Best Local Similarity 49.5%; Pred. No. 0.021;  
Matches 145; Conservative 0; Mismatches 144; Indels 4; Gaps 1;

QY 89 TGGTCTCAGGCTCTGCTCTGCTCAGGCGCGGCGAGCTGGGCTCGGCCACCTCTCCCTGGCG 247  
Db 306 TGTGCGCCCTTCTTGACCGCGCGAGCTGGGCTCGGCCACCTCTCCCTGGCG 247  
QY 149 CTGTTCTCAGGCTCTCAGC----CTGCTGATTTGCTTCTGCTGGGAGAAATGAG 204  
Db 246 GAGTTTCAGGAGACCTTCCGGCGCTCCGGGCTTGGCTTCGAGGGCGGAGAGGA 187  
QY 205 GTGGGAGAAACCCAGGCGAGTGTGTTGGAGTGAAGAGACAGCGCGGTGGGA 264  
Db 186 GCGTGAAG 127  
QY 265 GGTGAGAGAGAAATCTGCTGGGCTGGGAGTGGTGTGGGATCAACTGTCCCATTTGCTGCA 324  
Db 126 GGGCGCAGATCGCGGAGAGTGTGACGGGGCGGCGGAGAGAGAGAGAGAGAG 67  
QY 325 GGCTGTCTTGGGCGAGGAGAGGAGTGGGGGCCATAGCAGTGTGCTGCTCAGC 377

Db 66 CGGACAGACTGGCGGGTGGCGTGGCGGCGGAGTCCAGTGTGCGCGC 14

## RESULT 6

AAI81967/c

ID AAI81967 standard; cDNA; 357 BP.

XX AAJ81967;

AC 06-NOV-2001 (first entry)

DT Human polynucleotide SEQ ID NO 2027.

DE Human; cytokine; cell proliferation; cell differentiation; gene therapy;  
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;  
KW tissue growth factor; immunomodulatory; cancer; leukaemia;  
KW nervous system disorders; arthritis; inflammation; ss.

XX Homo sapiens.

XX WO200164835-A2.

PN 07-SEP-2001.

PD 26-FEB-2001; 2001WO-US04927.

PF 28-FEB-2000; 2000US-0515126.

PR 18-MAY-2000; 2000US-0577409.

XX (HYSE-) HYSEQ INC.

XX Tang YT, Liu C, Drmanac RT;

XX WPI; 2001-514838/56.

DR P-PSDB; AAO2036.

XX Isolated nucleic acids and polypeptides, useful for preventing  
PT diagnosing and treating e.g. leukaemia, inflammation and immune  
PT disorders.

XX Claim 1; SEQ ID NO 2027; 1399pp + Sequence Listing; English.

XX The invention relates to human polynucleotides (AAI79941-AAI93841) and  
CC the encoded proteins (AAO00010-AAO13910) that exhibit activity elating to  
CC cytokine, cell proliferation or cell differentiation or which may induce  
CC production of other cytokines in other cell populations. The  
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or  
CC peptide therapy. The polypeptides have various cytokine-like activities,  
CC e.g. stem cell growth factor activity, haematopoiesis regulating  
CC activity, tissue growth factor activity, immunomodulatory activity and  
CC activin/inhibin activity and may be useful in the diagnosis and/or  
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and  
CC inflammation.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 357 BP; 56 A; 218 C; 22 G; 61 T; 0 other;

Query Match 4.3%; Score 42.6; DB 22; Length 357;

Best Local Similarity 48.9%; Pred. NO. 0.065;

Matches 114; Conservative 0; Mismatches 119; Indels 0; Gaps 0;

Oy 193 GGGGAGTAATAGGTGGGAGAAAACAGGCCAGGCGAGTTGGTGTGAGTGAAGACAGA 252

Db 342 GGGGGGGGAGGTGGTGGGGGGGGGGTGGGGGGTGGAGGGGGGGGGGGGAGAG 283

Oy 253 CGCGGTGGGGAGGTCAGGAGAAATCTGCTGGGCTGGGGATGGTGTGGGCATCAACTGT 312

Db 282 GGGGGGTGGAGGGGGGAGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 223

Oy 313 CCCATGCTGACGGCTGCTTCTTGGGCGAGGGAAGGGGATGGGGGGCCATAGCAGTCTGG 372

Db 222 GGGGGGGGAGGCGGG 163

Oy 373 TCAGCCAGGCTGGCCTGGGAAGTGGTGGCCAGCAGCTACTAAGAGCCAGGAAA 425

Db 162 GAGGGGGGGGGGGTGGGAGAGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 110

## RESULT 7

AAH68527

ID AAH68527 standard; DNA; 349980 BP.

XX AC AAH68527;

XX DT 26-SEP-2001 (first entry)

DE C glutamicum coding sequence fragment SEQ ID NO: 7062.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;  
KW organic acid synthesis; ds.

XX Corynebacterium glutamicum.

XX EP1108790-A2.

XX PD 20-JUN-2001.

XX PF 18-DEC-2000; 2000EP-0127688.

XX PR 16-DEC-1999; 99JP-0377484.

XX PR 07-APR-2000; 2000JP-0159162.

XX PR 03-AUG-2000; 2000JP-0280988.

XX (KYOW ) KYOWA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;

PI Tateishi N, Senoh A, Ikeda M, Ozaki A;

XX WPI; 2001-376931/40.

XX Novel polynucleotides derived from Coryneform bacteria, for identifying  
PT mutation point of a gene, measuring expression of a gene, analysing  
PT expression profile or pattern of a gene and identifying homologous gene

XX Disclosure; SEQ ID NO: 7062; 246pp + Sequence Listing; English.

XX The present invention provides a number of nucleotide and protein  
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These  
CC are useful for identifying the mutation point of a gene derived from a  
CC mutant of coryneform bacterium, measuring expression amount and  
CC analysing the expression profile or expression pattern of a gene derived  
CC from Coryneform bacterium, and identifying a homologue of a gene derived  
CC from coryneform bacterium. Coryneform bacteria are useful for producing  
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,  
CC particularly L-lysine. The present sequence is a nucleic acid described  
CC in the exemplification of the invention.

CC Note: The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic format directly from the  
CC European Patent Office.

XX Sequence 349980 BP; 79725 A; 90426 C; 98918 G; 80911 T; 0 other;

Query Match 4.3%; Score 42.6; DB 22; Length 349980;

Best Local Similarity 50.2%; Pred. NO. 0.99;

Matches 105; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

Oy 555 GCCCAATTTTAGGGCCATGTGATTTGGGGCTATGTGACTCATCTGTGAAGTGTGG 614

Db 187448 GCATCACTTCAGGATCCCGGCATGTGCGCCATGAGATCATGACGAGCGCGCTGC 187507

Oy 615 GCCAGGAGCTGTGGGACCTTTAAATGCCAGCCAGTCTCATGTGCCGAGCTTTGGGGTAG 674

Db 187508 GCGCGAAGCGCGCGTGGTATTGGCGGAGTCACTGTTTCAGCGTACTGTCGGGGAG 187567

QY 675 GCCTAGGTAGGATGTGGATAATGGGAGGAGGAGGAGGATCTGTCTACCTAGGAGGCATC 734

Db 187568 GCCTCGGTCCAAAGTGTGCTGAGGTAGAGGTGGTATCCGCGGGAACCTCGTGGAGTT 187627

QY 735 CTCATCCATCCTTGGCCCGCGGACAGAGA 763

Db 187628 CAAGTTCAACGTGTGACGGCGCGAGAAGA 187656

RESULT 8

ABAL7086

ID ABAL7086 standard; DNA; 30032 BP.

XX AC

XX AC

XX ABAL7086;

DT 23-JAN-2002 (first entry)

XX DE Human nervous system related polynucleotide SEQ ID NO 9417.

XX DE Human; neurotropic; neuroprotective; cytostatic; dermatological; virucide;

XX KW immunosuppressive; antiinflammatory; anti-HIV; antibacterial; vulnerary;

XX KW antiparkinsonian; antisickling; antianaemic; antiarthritic; cancer;

XX KW antirheumatic; hepatotropic; cerebroprotective; antiinflammatory;

XX KW antiallergic; antidiabetic; antitumor; anticonvulsant; antifungal;

XX KW antiparasitic; cardiant; immune disorder; cardiovascular disorder;

XX KW neurological disease; infection; nephrotropic; gene therapy; vaccine; ds.

XX OS Homo sapiens.

XX PN WO200159063-A2.

XX PD 16-AUG-2001.

XX PF 17-JAN-2001; 2001WO-US01334.

XX PR 31-JAN-2000; 2000US-0179065.

XX PR 04-FEB-2000; 2000US-0180628.

XX PR 24-FEB-2000; 2000US-0184664.

XX PR 02-MAR-2000; 2000US-0186350.

XX PR 16-MAR-2000; 2000US-0189874.

XX PR 17-MAR-2000; 2000US-0190076.

XX PR 18-APR-2000; 2000US-0198123.

XX PR 19-MAY-2000; 2000US-0205515.

XX PR 07-JUN-2000; 2000US-0209467.

XX PR 28-JUN-2000; 2000US-0214886.

XX PR 30-JUN-2000; 2000US-0215135.

XX PR 07-JUL-2000; 2000US-0215647.

XX PR 07-JUL-2000; 2000US-0216880.

XX PR 11-JUL-2000; 2000US-0217487.

XX PR 11-JUL-2000; 2000US-0217496.

XX PR 14-JUL-2000; 2000US-0218290.

XX PR 26-JUL-2000; 2000US-0220963.

XX PR 26-JUL-2000; 2000US-0220964.

XX PR 14-AUG-2000; 2000US-0224518.

XX PR 14-AUG-2000; 2000US-0224519.

XX PR 14-AUG-2000; 2000US-0225213.

XX PR 14-AUG-2000; 2000US-0225214.

XX PR 14-AUG-2000; 2000US-0225266.

XX PR 14-AUG-2000; 2000US-0225267.

XX PR 14-AUG-2000; 2000US-0225268.

XX PR 14-AUG-2000; 2000US-0225270.

XX PR 14-AUG-2000; 2000US-0225447.

XX PR 14-AUG-2000; 2000US-0225757.

XX PR 14-AUG-2000; 2000US-0225758.

XX PR 14-AUG-2000; 2000US-0225759.

XX PR 18-AUG-2000; 2000US-0226279.

XX PR 22-AUG-2000; 2000US-0226681.

XX PR 22-AUG-2000; 2000US-0226868.

XX PR 22-AUG-2000; 2000US-0227182.

XX PR 23-AUG-2000; 2000US-0227009.

XX PR 30-AUG-2000; 2000US-0228924.

PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 03-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0234998.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 20-OCT-2000; 2000US-0242221.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.

PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249220.  
PR 17-NOV-2000; 2000US-0249221.  
PR 17-NOV-2000; 2000US-0249222.  
PR 17-NOV-2000; 2000US-0249223.  
PR 17-NOV-2000; 2000US-0249224.  
PR 17-NOV-2000; 2000US-0249225.  
PR 17-NOV-2000; 2000US-0249226.  
PR 17-NOV-2000; 2000US-0249227.  
PR 17-NOV-2000; 2000US-0249228.  
PR 17-NOV-2000; 2000US-0249229.  
PR 17-NOV-2000; 2000US-0249230.  
PR 01-DEC-2000; 2000US-0250391.  
PR 01-DEC-2000; 2000US-0251160.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HOMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX WPI; 2001-541565/60.  
XX  
XX Nucleic acids encoding 3224 human nervous system antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating nervous system  
PT cancers and metastases -  
XX  
XX Disclosure; SEQ ID NO 9417; 1701pp + Sequence Listing; English.  
XX  
XX The invention relates to novel genes (AB11004-AB21534) and proteins  
CC (AB114678-AB18001) useful for preventing, treating or ameliorating  
CC medical conditions e.g. by protein or gene therapy. The genes are  
CC isolated from a range of human tissues disclosed in the specification.  
CC The nucleic acids, proteins, antibodies and (antagonists are useful  
CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast  
CC and ovarian cancer and other cancers of the adrenal gland, bone, bone  
CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;  
CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune  
CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's  
CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative  
CC colitis; (c) cardiovascular disorders such as myocardial ischaemias;  
CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and  
CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal  
CC and parasitic infections.  
CC Note: The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 30032 BP; 6224 A; 9088 C; 9280 G; 5440 T; 0 other;  
SQ  
Query Match 4.2%; Score 42.2; DB 22; Length 30032;  
Best Local Similarity 53.3%; Pred. No. 0.49;  
Matches 89; Conservative 0; Mismatches 78; Indels 0; Gaps 0;  
QY 196 GAGAAATGAGTGGGAGAAACCCAGGCCAGTGTGGTGTGGAGTGAAGACGACAGCG 255  
DB 10048 GAGGGGATGTGGGGACAGAGGGGAGTGAAGACAGAGGGGTGTGGGGACAGAGG 10107  
QY 256 CGGTGGGAGGTGAGGAGAGAAATCTGTCGGCTGGGATGTGGGCATCACTGTCCC 315  
DB 10108 CGGTGGGGAACAGAGGGGGAGTGGGGACAGAGAGTATGTGGGGACAGAGGGGT 10167

QY 316 ATTGCTCAGGCTGTGTTGGGACAGGAGGAGGGATGGGGGCCATA 362  
DB 10168 GTGGGGACAGAGGGGGTGTGGAACACAGGAGGTGTGGGGGACAGA 10214  
RESULT 9  
AAK83781/c  
ID AAK83781 standard; DNA; 37664 BP.  
XX  
XX AAK83781;  
XX  
XX 07-NOV-2001 (first entry)  
XX Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38593.  
XX Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
XX cytostatic; gene therapy; vaccine; metastasis; ds.  
XX Homo sapiens.  
XX WO200157182-A2.  
XX  
XX 09-AUG-2001.  
XX  
XX 17-JAN-2001; 2001WO-US01354.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225266.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226688.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.





XX AAK83782; 14-SEP-2000; 2000US-0232400.  
AC 14-SEP-2000; 2000US-0232401.  
XX 14-SEP-2000; 2000US-0233063.  
DT 14-SEP-2000; 2000US-0233064.  
XX 14-SEP-2000; 2000US-0233065.  
DE 21-SEP-2000; 2000US-0234223.  
XX 21-SEP-2000; 2000US-0234274.  
KW 25-SEP-2000; 2000US-0234997.  
XX 25-SEP-2000; 2000US-0234998.  
KW 26-SEP-2000; 2000US-0235484.  
XX 27-SEP-2000; 2000US-0235834.  
OS 27-SEP-2000; 2000US-0235836.  
PN 29-SEP-2000; 2000US-0236327.  
XX 29-SEP-2000; 2000US-0236367.  
PD 29-SEP-2000; 2000US-0236368.  
XX 29-SEP-2000; 2000US-0236369.  
XX 29-SEP-2000; 2000US-0236370.  
XX 02-OCT-2000; 2000US-0236802.  
XX 02-OCT-2000; 2000US-0237037.  
XX 02-OCT-2000; 2000US-0237038.  
XX 02-OCT-2000; 2000US-0237039.  
XX 02-OCT-2000; 2000US-0237040.  
XX 13-OCT-2000; 2000US-0239935.  
XX 13-OCT-2000; 2000US-0239937.  
XX 20-OCT-2000; 2000US-0240960.  
XX 20-OCT-2000; 2000US-0241221.  
XX 20-OCT-2000; 2000US-0241785.  
XX 20-OCT-2000; 2000US-0241786.  
XX 20-OCT-2000; 2000US-0241787.  
XX 20-OCT-2000; 2000US-0241808.  
XX 20-OCT-2000; 2000US-0241809.  
XX 20-OCT-2000; 2000US-0241826.  
XX 01-NOV-2000; 2000US-0244617.  
XX 08-NOV-2000; 2000US-0246474.  
XX 08-NOV-2000; 2000US-0246475.  
XX 08-NOV-2000; 2000US-0246476.  
XX 08-NOV-2000; 2000US-0246477.  
XX 08-NOV-2000; 2000US-0246478.  
XX 08-NOV-2000; 2000US-0246523.  
XX 08-NOV-2000; 2000US-0246524.  
XX 08-NOV-2000; 2000US-0246525.  
XX 08-NOV-2000; 2000US-0246526.  
XX 08-NOV-2000; 2000US-0246527.  
XX 08-NOV-2000; 2000US-0246528.  
XX 08-NOV-2000; 2000US-0246532.  
XX 08-NOV-2000; 2000US-0246609.  
XX 08-NOV-2000; 2000US-0246610.  
XX 08-NOV-2000; 2000US-0246611.  
XX 08-NOV-2000; 2000US-0246613.  
XX 17-NOV-2000; 2000US-0249207.  
XX 17-NOV-2000; 2000US-0249208.  
XX 17-NOV-2000; 2000US-0249209.  
XX 17-NOV-2000; 2000US-0249210.  
XX 17-NOV-2000; 2000US-0249211.  
XX 17-NOV-2000; 2000US-0249212.  
XX 17-NOV-2000; 2000US-0249213.  
XX 17-NOV-2000; 2000US-0249214.  
XX 17-NOV-2000; 2000US-0249215.  
XX 17-NOV-2000; 2000US-0249216.  
XX 17-NOV-2000; 2000US-0249217.  
XX 17-NOV-2000; 2000US-0249218.  
XX 17-NOV-2000; 2000US-0249244.  
XX 17-NOV-2000; 2000US-0249245.  
XX 17-NOV-2000; 2000US-0249264.  
XX 17-NOV-2000; 2000US-0249265.  
XX 17-NOV-2000; 2000US-0249297.  
XX 17-NOV-2000; 2000US-0249299.  
XX 01-DEC-2000; 2000US-0250160.  
XX 01-DEC-2000; 2000US-0250391.  
XX 05-DEC-2000; 2000US-0251030.  
XX 05-DEC-2000; 2000US-0251988.  
XX 05-DEC-2000; 2000US-0256719.  
  
Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38594.  
Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;  
cytostatic; gene therapy; vaccine; metastasis; ds.  
Homo sapiens.  
WO200157182-A2.  
09-AUG-2001.  
17-JAN-2001; 2001WO-US01354.  
31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225279.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.

PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI: 2001-483426/52.  
XX  
XX Nucleic acids encoding human immune/hematopoietic antigen polypeptides,  
PT useful for preventing, diagnosing and/or treating cancers and  
PT metastasis -  
XX  
XX Disclosure; SEQ ID NO 38594; 3071pp + Sequence Listing; English.  
XX  
XX AAK54951 to AAK64702 encode the human immune/hematopoietic antigen (I)  
CC amino acid sequences given in AAK82170 to AAK91921. (I) have cytostatic  
CC activity, and can be used in gene therapy and vaccine production. (I)  
CC proteins and polynucleotides may be used in the prevention, diagnosis and  
CC treatment of diseases associated with inappropriate (I) expression. For  
CC example, they may be used to treat disorders associated with decreased  
CC expression by rectifying mutations or deletions in a patient's genome  
CC that affect the activity of (I) by expressing inactive proteins or to  
CC supplement the patients own production of (I). Additionally, (I)  
CC polynucleotides may be used to produce the secreted (I), by inserting  
CC the nucleic acids into a host cell and culturing the cell to express the  
CC protein. (I) proteins and polynucleotides may be used to prevent,  
CC diagnose and treat immune/hematopoietic-related diseases, especially  
CC cancers and cancer metastases of hematopoietic-derived cells. AAK64703  
CC to AAK87694 represent human immune/hematopoietic antigen genomic  
CC sequences from the present invention. AAK54942 to AAK54950 and AAK82169  
CC represent sequences used in the exemplification of the present invention.  
XX  
SQ Sequence 61710 BP; 11560 A; 18542 C; 18148 G; 13460 T; 0 other;  
Query Match 4.2%; Score 42; DB 22; Length 61710;  
Best Local Similarity 50.0%; Pred. No. 0.73;  
Matches 105; Conservative 0; Mismatches 105; Indels 0; Gaps 0;  
QY 195 GGAGATGAGGTGGGAGAAACAGCCAGGCGAGTGTGTGGAGTGAAGACGACG 254  
DB 15325 GGTGATGAGCAGCAGCCAGGCGAGGCGCCCTGGAGTTGAGTGGCAGTGGCAGC 15266  
QY 255 GCGGTGGGAGGTGAGGAGAGATCTGTGGGCTGGGGATGGTGGGCAATCACTGTCC 314  
DB 15265 GGGCCCTGGAGGTGAGTACAGCTGGGTATGCCGGAGGGCCCTGGAGGTGAGTGGCAA 15206  
QY 315 CATTCGTGAGCTGTCTTGGGCGAGGAGGAGTGGGGCCCATGACAGTGTGGTC 374  
DB 15205 TTGGGGGAGGGGGCCCTGGAGGCGGGCAGGCCCTGGAGGTGAATAGCAGTGGGTAT 15146  
QY 375 AGCCAGGCTGGCTGGGAAGTGTGCTCCAG 404  
DB 15145 GCCCGGCGAGGCCCTGGAGGTGAGTGGCAG 15116  
RESULT 11  
ID ABK10687/c  
XX ABK10687 standard; DNA; 9359 BP.  
AC ABK10687;  
XX  
XX 05-JUN-2002 (first entry)  
XX Transformation vector plasmid pVDH636 DNA.  
DE Grass; plant; transgenic; flowering inhibition; inflorescence; gene; ds;  
XX

KW tiller production; delayed heading; gibberellic acid; phytohormone; golf;  
KW genetically modified grass; athletic field; sport; baseball; cricket;  
KW football; rugby; soccer; tennis; lawn; landscaping; cattle; horse; sheep;  
KW goat; animal feedstuff; Arabidopsis thaliana; Ath1; homeobox gene;  
XX circular; cyclic; pVDH636.  
OS Synthetic.  
XX

PH Key Location/Qualifiers  
FT CDS 4930..6351  
FT /\*tag= a  
FT /product= "Arabidopsis thaliana Ath1"

XX WO200214486-A2.

XX 21-FEB-2002.

XX 16-AUG-2001; 2001WO-EP09570.

XX 18-AUG-2000; 2000US-226422P.

XX 27-NOV-2000; 2000US-253274P.

XX 22-JUN-2001; 2001US-300220P.

XX (ADVA-) ADVANTA SEEDS BV.

XX Van Der Valk P, Van Dun CMP, Smeekens SCM, Proveniers MCG;

XX WPI: 2002-257603/30.

XX P-PSDB; AAU76885.

XX New genetically modified grass useful for growing and/or propagating  
PT grass in athletic fields (for sports such as baseball, cricket,  
PT football, golf, rugby, soccer and tennis), or as animal feedstuff for  
PT cattle, goat, horse and sheep -  
XX

XX Example 1; Fig 6; 45pp; English.

CC The invention relates to a grass plant which has been genetically  
CC modified to substantially inhibit generative propagation. The genetic  
CC modification may result in a heritable change in one or more plant  
CC characteristics such as inhibition of flowering (or substantial delay  
CC that amounts to inhibition), absence of inflorescence, increased  
CC production of tillers, delayed heading and inhibition of the  
CC developmental switch from vegetative to generative growth. A method of  
CC making a grass involves transforming the grass with a nucleic acid which  
CC interferes with metabolism of gibberellic acid. A grass can be treated by  
CC applying a phytohormone to at least partially relieve or reverse a change  
CC in plant characteristic resulting from genetic modification. The  
CC genetically modified grass is useful for growing and/or propagating grass  
CC in athletic fields (for sports such as baseball, cricket, football, golf,  
CC rugby, soccer and tennis), lawns, parks and other types of landscaping.  
CC The grass is also useful as an animal feedstuff for cattle, goats, horses  
CC and sheep, due to its increased vegetative growth, improved digestibility  
CC and/or nutritional value as animal feedstuff. This sequence represents  
CC plasmid pVDH636 DNA. This plasmid was used as a transformation vector for  
CC production of transgenic grasses expressing the Arabidopsis thaliana  
CC homeobox gene, Ath1.

XX Sequence 9359 BP; 2377 A; 2222 C; 2133 G; 2624 T; 3 other;

Query Match 4.1%; Score 41.4; DB 24; Length 9359;  
Best Local Similarity 50.6%; Pred. No. 0.51;  
Matches 119; Conservative 3; Mismatches 112; Indels 1; Gaps 1;

QY 119 CCAGGGCCCCCAACCCGCCGCCAGCAGCTGTTCTACAGCCCTCTCAGCTGTCTGATT 178  
DB 7416 CCGGCCGCCACAGACCCGCCGCCACCTGCTGTTTTTCTTTTCTTTTCTTTCTTTT 7357

QY 179 TGCTTGTCTGCTGGGGAGAAATGAGTGGGAGAAACAGCCAGCCAGCTTGGTGTG 238  
DB 7356 TTTTCTTTTGGTGGGAGACGTGCGGTGCGGCAACTCAGGTGATAGTGGGGG 7297

Qy	239	GAGTGAAGACCAACGGCGGTGGGAGGTCAGAGAGAAATCTGCTGGGCTGGGATGGT	298
Db	7296	GHVTVGAGACATATTGCCAGTTGGCTGGACTGGGTGGTGGTGGTGGTGGTGGTGGC	7237
Qy	299	TGGGCATCAACTGTCCCATGTCT-GCAGCGCTGGTCTTGGGCGAGGGAAGGGGATG	352
Db	7236	TGGGVTGTGATGGATCGTGGATAGCACATTTGGCGCTTTAGGAACATTTAGGGGTG	7182

```

RESULT 12
ABK11039/c
ID   ABK11039 standard; DNA: 9359 BP.
XX
XX   ABK11039;
XX   AC
XX   AC
XX   DT
XX   DT
XX   05-JUN-2002 (first entry)
XX
XX   pVDH636 vector containing Arabidopsis Ath1 gene.
XX

```

PN	WO200214524-A2.	
XX		
PD	21-FEB-2002.	
XX		
XX	16-AUG-2001; 2001WO-EP09572.	
XX		
PR	18-AUG-2000; 2000US-226422P.	
PR	27-NOV-2000; 2000US-253274P.	
PR	27-NOV-2000; 2000US-253327P.	
PR	22-JUN-2001; 2001US-300220P.	
XX		
XX	(ADVA-) ADVANTA SEEDS BV.	
XX		
XX	Van Der Valk P, Van Dun CMP, Smeekens SCM, Proveniers MCG;	
PI		
XX	WPI; 2002-257611/30.	
DR	P-PSDB; AAU76514.	
DR		
XX		
PT	New genetically modified grasses that express inhibited generative	
PT	propagation, or herbicide resistance, useful for forage (e.g. cattle	
PT	feedstuff) or amenity purposes (e.g. for use in an athletic field, lawn	
PT	or park) -	
XX		
XX	Example 1; Fig 6; 56pp; English.	
PS		
XX		
CC	The invention relates to a grass plant, which has been genetically	
CC	modified to substantially inhibit generative propagation and carry	
CC	herbicide resistance. The grass is useful for growth and/or propagation	
CC	of grasses. The grass is used to play at least one sport (e.g. baseball,	
CC	cricket, football, golf, rugby, soccer or tennis), or used at least in a	
CC	portion of an athletic field, lawn or park. The grass is fed to animal	
CC	(e.g. cattle, goat, horse or sheep) or used as an animal feedstuff.	
CC	The present sequence represents the coding sequence of pVDH636 vector	
CC	containing the Arabidopsis Ath1 gene, which was used to inhibit flowering	

CC in grasses.

XX  
SQ Sequence 9359 BP; 2377 A; 2222 C; 2133 G; 2624 T; 3 other;

Query Match	4.1%;	Score 41.4;	DB 24;	Length 9359;
Best Local Similarity	50.6%;	Pred. No. 0.51;		
Matches 119; Conservative	3;	Mismatches 112;	Indels 1;	

Qy	119	CGAGGCCGCCAACCCGCCGCCCGCCAGCAGCTGTTCTCAGGCCCTCAGCCTGCTCGT	178
Db	7416	CCGGCCCCCAGCAGCCCGGACCACTGCTGTTTTTCTTTTTTCTTTTTTCTTTCTTTT	7357
Qy	179	TGCTTCTCGCCTGGGAGAAATGAGGTGGGAGAAAACAGGCCAGGCCAGTTGCTGTG	238
Db	7356	TTTTTTTTTGGCTGGGAGACGTGCGGTGCGTGCAGCACTACAGGTGATGTGGGGGG	7297
Qy	239	GAGTGAAGAGCAGACGGCGGTGGGAGGTCAGGAGAGAAATCTGCTGGCTCGGGATG	298
		:::	
Db	7296	GHVTGGAGACTATTGCCAGTTGGCTGGACATGGGCTGGGTGGGTTGGGTGGCTGGC	7237
Qy	299	TGGGCATCACTGTCCATTGCT-GCAGGCTGGTCTTGGGCGAGGGAAGGGGATG	352
Db	7236	TGGGVTTCGATCGAGTCGTGGATAGACATTTGGGCTTTAGGAACATTTAGGGGTTG	7182

RESULT 13	
AAQ20466/c	
ID	AAQ20466 standard; DNA; 1894 BP.
XX	
XX	AAQ20466;
XX	
XX	
DT	14-APR-1992 (first entry)
XX	
DE	Sequence of chromosomal gene, including promoter, for porcine
DE	C-type natriuretic peptide (CPN) precursor.
XX	
XX	
KW	Hypotensive agent; natriuretic; brain-specific promoter; ss.
XX	
OS	Sus scrofa domestica.

Key	Location/Qualifiers
promoter	80...84
	/*tag= a
	/label= Y box
promoter	88...93
	/*tag= b
	/label= GC box
promoter	100...105
	/*tag= c
	/note= "GC box"
TATA_signal	133...138
	/*tag= d
exon	310...399
	/*tag= e
exon	839...1229
	/*tag= f

XX	EP466175-A.	
XX		
XX	15-JAN-1992.	
XX		
XX	12-JUL-1991; 91EP-0111630.	
XX		
XX	13-JUL-1990; 90JP-0186583.	
XX		
XX	(SUNR ) SUNTORY LTD.	
XX		
XX	Matsuo H, Kangawa K, Minam	
XX	Tawaragi Y;	
XX		
XX	WPI; 1992-017806/03.	
XX	P-PSDB; RAR20074.	
XX		



AC AAS29976;  
XX 21-NOV-2001 (first entry)  
XX Human lung antigen genomic DNA #46.  
DE  
XX  
XX Lung antigen protein; human; mouse; rabbit; goat; horse; cat; dog;  
KW chicken; sheep; immunosuppressive; antiarthritic; vasotropic;  
KW antirheumatic; antiproliferative; cytostatic; cardiant; neuroprotective;  
KW cerebroprotective; nootropic; antibacterial; virucide; fungicide; cancer;  
KW ophthalmological; vulnerary; gene therapy; autoimmune disease; neoplasm;  
KW hyperproliferative disorder; breast; liver; cardiovascular disorder; ds;  
KW cerebrovascular disorder; nervous system disorder; bacterial infection;  
KW fungal infection; viral infection; ocular disorder; endocrine disorder;  
KW gastrointestinal disorder; renal disorder; respiratory disorder;  
KW wound healing; skin aging; organ transplantation; food preservative;  
KW tissue regeneration; anti-infertility; food additive.  
XX  
OS Homo sapiens.  
XX  
XX WO200155303-A2.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001WO-US01301.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0232968.  
PR 14-SEP-2000; 2000US-0232997.  
PR 14-SEP-2000; 2000US-0232998.  
PR 14-SEP-2000; 2000US-0232999.  
PR 14-SEP-2000; 2000US-0233400.  
PR 14-SEP-2000; 2000US-0233401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.

PR	17-NOV-2000;	2000US-0249297.
PR	17-NOV-2000;	2000US-0249299.
PR	17-NOV-2000;	2000US-0249300.
PR	01-DEC-2000;	2000US-0250160.
PR	01-DEC-2000;	2000US-0250391.
PR	05-DEC-2000;	2000US-0251030.
PR	03-DEC-2000;	2000US-0251988.
PR	05-DEC-2000;	2000US-0256719.
PR	06-DEC-2000;	2000US-0251479.
PR	08-DEC-2000;	2000US-0251856.
PR	08-DEC-2000;	2000US-0251868.
PR	08-DEC-2000;	2000US-0251869.
PR	08-DEC-2000;	2000US-0251989.
PR	08-DEC-2000;	2000US-0251990.
PR	11-DEC-2000;	2000US-0254097.
PR	05-JAN-2001;	2001US-0259678.
XX		
XX	(HUMA-) HUMAN GENOME SCI INC.	
XX		
XX	Rosen CA, Barash SC, Ruben SM;	
XX	WPI; 2001-457723/49.	
XX		
PT	Isolated polypeptide for treating, preventing and/or prognosing	
PT	respiratory disorders related to the lung including lung cancers and	
PT	also for testing and detection e.g. diagnosis -	
XX		
XX	Claim 1; SEQ ID No 240; 507pp; English.	
XX		
CC	Sequences AAS2931-AA330164 represent genomic DNA molecules, which encode	
CC	the lung antigen polypeptides of the invention. Lung antigen polypeptides	
CC	and their associated polynucleotides are useful in the diagnosis,	
CC	treatment and prevention of various types of disorders in e.g. humans,	
CC	mice, rabbits, goats, horses, cats, dogs, chickens or sheep. A	
CC	pathological condition can be determined by detecting the presence or	
CC	absence of a mutation in a lung antigen polynucleotide. The treatable	
CC	disorders include autoimmune diseases such as rheumatoid arthritis,	
CC	hyperproliferative disorders such as neoplasms of the breast or liver,	
CC	cardiovascular disorders such as cardiac arrest, cerebrovascular	
CC	disorders such as cerebral ischemia, nervous system disorders such as	
CC	Alzheimer's disease, infections caused by bacteria, viruses and fungi,	
CC	ocular disorders such as corneal infection, endocrine disorders such as	
CC	premature labour and infertility, gastrointestinal disorders such as	
CC	Crohn's disease, renal disorders such as glomerulonephritis and	
CC	respiratory disorders such as asthma and pleurisy. The polypeptides can	
CC	also be used to aid wound healing, to prevent skin aging due to sunburn,	
CC	to maintain organs before transplantation, to regenerate tissues and in	
CC	chemotaxis. The polypeptides can also be used as a food additive or	
CC	preservative to increase or decrease storage capabilities.	
CC	Note: The sequence data for this patent did not form part of the printed	
CC	specification, but was obtained in electronic format directly from WIPO	
CC	at <a href="http://ftp.wipo.int/pub/published_pct_sequences">ftp.wipo.int/pub/published_pct_sequences</a> .	





```
Db 194 LFLGQTGLTTEEMGKALEPGENFCMGGPGVIMSRVLRVPHIGKCLREMYTHDEV 253
QY 522 AGCCCTGGTGTCCCATGCTCTCTCTGTGCGCCCAATTTTAGGCCCATGTGATTGG 581
Db 254 EVGRCVRRFAGVQCVMSYEMRQLFYENYEQNKKGYIRDLHNSKIHQAITLHPNKNPPYQY 313
QY 582 GGGCTATGCTACTGCTCTCTAAGTGCTTGGGCCAGGAGCTGTGGCCACCTTTAAATG 641
Db 314 RLHYSMLSRKISELRHRIQIHLREIVLMSKYSNTEIHKEDLQLGIPPFMRFPQOREI 373
QY 642 CCAGCCAGTCATGCTGCCGAGTTTGGGTTAGGCTAGGTAGGATTTGGAATATGGGA 701
Db 374 LEWEFLTKYLYSAVDGPPRRGMDSAQREALDDIVMQVMEMINAKTRIGRIIDFKEIQ 433
QY 702 GGAGCGAGGATCTGTCTACCTAGGAGGCATCTCATCCATCCCTGGCCCTGGACAGA 761
Db 434 YGYRRVPMWGAEYILDLLLYLKKHKKMTVPVRRHAYLQOTFSKIQFVEHEELDAOEL 493
QY 762 GAACCTGAACCTTGGTAGGGCCTCAGGACGATGCTGCGTGGCCCTTGGGAATCTGGA 821
Db 494 AKRINQESGSLFSLNSLKLVPQLPGSKSEHKPKDKKINILPLSGRDFMVRPMGN 553
QY 822 TTGTCCTGGTCATAGTTCTTATCTGTGACCAACACCCCTTAGCTGCCAGGCTTTGGACA 881
Db 554 FEKTCLIPNQWKLVLVLLFNSDSNPKAKQVELMRDYRIKYPKADMQILPVSGEFSRALA 613
QY 882 TGG 884
Db 614 LEV 616
```

## RESULT 2

```
US-10-079-854-240/C
; Sequence 240, Application US/10079854
; Publication No. US20030034368A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121C1
; CURRENT APPLICATION NUMBER: US/10/079,854
; CURRENT FILING DATE: 2002-02-22
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 240
; LENGTH: 5786
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-10-079-854-240
```

```
Query Match 3.9%; Score 39.2; DB 9; Length 5786;
Best Local Similarity 51.1%; Pred. No. 0.029;
Matches 92; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 29 TTGCTATTGTTACTGCTTTTACGTCCTTGGAAAAAGTTAGCACAAAGGCTGCTTTG 88
Db 4527 TTTTCTTTGAGATAGGCTCTCAGTCTGTACCCAGGTGGAGTACAGTGGGTGATCTC 4468
QY 89 TGGCTACACCCCTCTGCTCTGCGCTCACCAGGCCCCCAACCCGCCCCCCCCCAGCAG 148
Db 4467 AGCTCCCTGCTCTCTGCTCTCCAGGTTCAACTGATTTCTCCACCTCAGCTCCAGAGCAG 4408
QY 149 CTGTTCTCAGGCTCTCAGCTGCTGATTTGCTGTGCGCTGGGGAGAAATGAGGTGG 208
Db 4407 CTGGGACCACGAGCATGCACCACTATGCCAGCTAATTTTGGTGGGTATTTTGGTAG 4348
```

## RESULT 3

```
US-09-764-878-240/c
; Sequence 240, Application US/09764878
; Patent No. US20020090615A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PA121
; CURRENT APPLICATION NUMBER: US/09/764,878
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 428
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 240
; LENGTH: 5786
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-764-878-240

Query Match 3.9%; Score 39.2; DB 10; Length 5786;
Best Local Similarity 51.1%; Pred. No. 0.029;
Matches 92; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

QY 29 TTGCTATTGTTACTGCTTTTACGTCCTTGGAAAAAGTTAGCACAAAGGCTGCTTTG 88
Db 4527 TTTTCTTTGAGATAGGCTCTCAGTCTGTACCCAGGTGGAGTACAGTGGGTGATCTC 4468
QY 89 TGGCTACACCCCTCTGCTCTGCGCTCACCAGGCCCCCAACCCGCCCCCCCCCAGCAG 148
Db 4467 AGCTCCCTGCTCTCTGCTCTCCAGGTTCAACTGATTTCTCCACCTCAGCTCCAGAGCAG 4408
QY 149 CTGTTCTCAGGCTCTCAGCTGCTGATTTGCTGTGCGCTGGGGAGAAATGAGGTGG 208
Db 4407 CTGGGACCACGAGCATGCACCACTATGCCAGCTAATTTTGGTGGGTATTTTGGTAG 4348
```

## RESULT 4

```
US-09-989-442-47/C
; Sequence 47, Application US/09989442
; Publication No. US20030013649A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: P4208
; CURRENT APPLICATION NUMBER: US/09/989,442
; CURRENT FILING DATE: 2001-11-21
; Prior Application Number: 60/179,065
; Prior Filing Date: 2000-01-31
; Prior Application Number: 60/180,628
; Prior Filing Date: 2000-02-04
; Prior Application Number: 60/214,886
; Prior Filing Date: 2000-06-28
; Prior Application Number: 60/217,487
; Prior Filing Date: 2000-07-11
; Prior Application Number: 60/225,758
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/220,963
; Prior Filing Date: 2000-07-26
; Prior Application Number: 60/217,496
; Prior Filing Date: 2000-07-11
; Prior Application Number: 60/225,447
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/218,290
; Prior Filing Date: 2000-07-14
; Prior Application Number: 60/225,757
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/226,868
; Prior Filing Date: 2000-08-22
; Prior Application Number: 60/216,647
; Prior Filing Date: 2000-07-07
; Prior Application Number: 60/225,267
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/216,880
; Prior Filing Date: 2000-07-07
; Prior Application Number: 60/225,270
; Prior Filing Date: 2000-08-14
; Prior Application Number: 60/251,869
; Prior Filing Date: 2000-12-08
```

; PRIOR APPLICATION NUMBER: 60/235,834  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/234,274  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/234,223  
; PRIOR FILING DATE: 2000-09-21  
; PRIOR APPLICATION NUMBER: 60/228,924  
; PRIOR FILING DATE: 2000-08-30  
; PRIOR APPLICATION NUMBER: 60/224,518  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,369  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/224,519  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/220,964  
; PRIOR FILING DATE: 2000-07-26  
; PRIOR APPLICATION NUMBER: 60/241,809  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/249,299  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/236,327  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/241,785  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/244,617  
; PRIOR FILING DATE: 2000-11-01  
; PRIOR APPLICATION NUMBER: 60/225,268  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/236,368  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/251,856  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/251,868  
; PRIOR FILING DATE: 2000-12-08  
; PRIOR APPLICATION NUMBER: 60/229,344  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/234,997  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: 60/229,343  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,345  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,287  
; PRIOR FILING DATE: 2000-09-01  
; PRIOR APPLICATION NUMBER: 60/229,513  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/231,413  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/229,509  
; PRIOR FILING DATE: 2000-09-05  
; PRIOR APPLICATION NUMBER: 60/236,367  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/237,039  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,038  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/236,370  
; PRIOR FILING DATE: 2000-09-29  
; PRIOR APPLICATION NUMBER: 60/236,802  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,037  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/237,040  
; PRIOR FILING DATE: 2000-10-02  
; PRIOR APPLICATION NUMBER: 60/240,960  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/239,935  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/239,937  
; PRIOR FILING DATE: 2000-10-13  
; PRIOR APPLICATION NUMBER: 60/241,787  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/246,474

; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/246,532  
; PRIOR FILING DATE: 2000-11-08  
; PRIOR APPLICATION NUMBER: 60/249,216  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,210  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/226,681  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,759  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/225,213  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/227,182  
; PRIOR FILING DATE: 2000-08-22  
; PRIOR APPLICATION NUMBER: 60/225,214  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/235,836  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: 60/230,438  
; PRIOR FILING DATE: 2000-09-06  
; PRIOR APPLICATION NUMBER: 60/215,135  
; PRIOR FILING DATE: 2000-06-30  
; PRIOR APPLICATION NUMBER: 60/225,266  
; PRIOR FILING DATE: 2000-08-14  
; PRIOR APPLICATION NUMBER: 60/249,218  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,208  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,213  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,212  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,207  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,245  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,244  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,217  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,211  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,215  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,264  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,214  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/249,297  
; PRIOR FILING DATE: 2000-11-17  
; PRIOR APPLICATION NUMBER: 60/232,400  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/231,242  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/232,081  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/232,080  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/231,414  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/231,244  
; PRIOR FILING DATE: 2000-09-08  
; PRIOR APPLICATION NUMBER: 60/233,064  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/233,063  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,397  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,399  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 60/232,401  
; PRIOR FILING DATE: 2000-09-14

;; PRIOR APPLICATION NUMBER: 60/241,808  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,826  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,786  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/241,221  
;; PRIOR FILING DATE: 2000-10-20  
;; PRIOR APPLICATION NUMBER: 60/246,475  
;; PRIOR FILING DATE: 2000-11-08  
;; PRIOR APPLICATION NUMBER: 60/231,243  
;; PRIOR FILING DATE: 2000-09-08  
;; PRIOR APPLICATION NUMBER: 60/233,065  
;; PRIOR FILING DATE: 2000-09-14

Query Match 3.8%; Score 38.4; DB 9; Length 2237;  
Best Local Similarity 51.18; Pred. No. 0.038; Mismatches 86; Indels 0; Gaps 0;  
Matches 90; Conservative 0;

QY 301 GGCATCACTGCTCCATTGCTGCAGGCTGTCTTGGGCGAGGGAAGGGGATGGGGGCCA 360  
||||| ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 955 GGCAGGCTCTCTCCATGCTGTCTCCAGCTGTCTCCAGCTGTCTCCAGCTCAAGTGAAGTGA 896  
QY 361 TAGCAGTGTGGTGCAGCAGCTGGCTGGGAAGTGGTGGCCAGGCATCTAAGAGCCA 420  
||||| ||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 895 TTATACATCTGGGCGAGAGTTGGGGGGCGGGGTGTGATCTCCAGAGCAGCTGTGCTGCCA 836  
QY 421 GGAAGCCCTGCCAAGTGTGTGGCTAGTTCCCTGTGCATCAGCGGCTAGCAGCC 476  
||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||  
Db 835 GGAATGACCTGCAGGCTCTCTGGCGCACAGGCAGCTGTCAGCCAGCTGCTGTC 780

## RESULT 5

US-09-783-590-7694/c  
; Sequence 7694, Application US/09783590  
; Patent No. US20020110850A1

## GENERAL INFORMATION:

;; APPLICANT: Idillon, Patrick J.  
;; APPLICANT: Haseltine, William A.  
;; APPLICANT: Li, Haodong  
;; APPLICANT: Rosen, Craig A.  
;; APPLICANT: Ruben, Steven M.  
;; TITLE OF INVENTION: Human Genes, Sequences, and Expression Products 16.2  
;; FILE REFERENCE: PO-16.2C1  
;; CURRENT APPLICATION NUMBER: US/09/783,590  
;; PRIOR FILING DATE: 2000-02-15  
;; PRIOR APPLICATION NUMBER: 08/420,856  
;; PRIOR FILING DATE: 1995-04-12  
;; PRIOR APPLICATION NUMBER: 08/346,731  
;; PRIOR FILING DATE: 1994-11-21  
;; NUMBER OF SEQ ID NOS: 12485  
;; SOFTWARE: PatentIn Ver. 2.0  
;; SEQ ID NO 7694  
;; LENGTH: 246  
;; TYPE: DNA  
;; ORGANISM: Homo sapiens  
;; FEATURE:  
;; NAME/KEY: misc feature  
;; LOCATION: (71)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (106)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (144)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (229)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature  
;; LOCATION: (242)  
;; OTHER INFORMATION: n equals a,t,g, or c  
;; NAME/KEY: misc feature

;; LOCATION: (245)  
;; OTHER INFORMATION: n equals a,t,g, or c  
US-09-783-590-7694

Query Match 3.8%; Score 38.2; DB 10; Length 246;  
Best Local Similarity 53.5%; Pred. No. 0.021; Mismatches 0; Gaps 0;  
Matches 76; Conservative 0;

QY 10 CTCCTCCCCACCACTGCTGGATTGCTCTATTCTTACTGCTTTTCTGCTTGTGAAAAAGTTAGC 69  
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 155 CTGCATGCCCANCTCGAGTTGCTTTTAACTTTGTTTATTTTATTTTAAAAAATG 96  
QY 70 ACAACAAGGGCTGCTTTGTGGCTCACCCCTCTGCTCTCTGCTCACCCAGGCCCCCCC 129  
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 95 CAGACAGGTCTTACTATGTTGCTNAGGCTCTTGAACCTCTGCTCAAGTATCCTCTCC 36  
QY 130 AACCCCGCCCCCCCCAGCAGCTG 151  
|-||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||  
Db 35 ATCCAGCGCTCCCGAGTGTGCTG 14

## RESULT 6

US-10-184-644-164/c  
; Sequence 164, Application US/10184644  
; Publication No. US20030044930A1  
; GENERAL INFORMATION:  
;; APPLICANT: Baker, Kevin P.  
;; APPLICANT: Chen, Jian  
;; APPLICANT: Desnoyers, Luc  
;; APPLICANT: Goddard, Audrey  
;; APPLICANT: Godowski, Paul J.  
;; APPLICANT: Gurney, Austin L.  
;; APPLICANT: Pan, James  
;; APPLICANT: Smith, Victoria  
;; APPLICANT: Watanabe, Colin K.  
;; APPLICANT: Wood, William I.  
;; APPLICANT: Zhang, Zemin  
;; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC  
;; TITLE OF INVENTION: ACIDS ENCODING THE SAME  
;; FILE REFERENCE: P3430R1C227  
;; CURRENT APPLICATION NUMBER: US/10/184,644  
;; CURRENT FILING DATE: 2002-06-28  
;; Prior Application removed - See File Wrapper or Palm  
;; NUMBER OF SEQ ID NOS: 612  
;; SEQ ID NO 164  
;; LENGTH: 299  
;; TYPE: PRT  
;; ORGANISM: Homo Sapien  
US-10-184-644-164

Query Match 3.8%; Score 38.2; DB 9; Length 299;  
Best Local Similarity 4.8%; Pred. No. 0.022; Mismatches 8; Conservative 99; Indels 0; Gaps 0;

QY 8 CCCTCCCCCACCAGCTGGATTGCTATTCTTACTGCTTTTACTGCTTTTACCTCTGGAAGAATTA 67  
|::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|  
Db 293 CSYSCDDMSMDSY..SYSY.YHYDDCY.DYDMTTHSDHYSMS.SYSYNNMYMDCDYN 234  
QY 68 GCACAACAAGGGCTGCTTTGTGGCTCACCCCTCTGCTCTGCTGCTCACCCAGGCCCC 127  
::: |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|  
Db 233 SSM.HHYADWTM.D.SCDNDYY.S.SDSYSYSYSRACYSYNNRDYI.ADSYSYS 174  
QY 128 CCAACCCCGCCCCCAGCAGCTGTTCTCAGGCTCTCAGCTGTC 173  
: |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::| |::|  
Db 173 YAYSYSYSTSYNNMYBNCNR.SMST..YH...MMBAMBN.S 128

## RESULT 7

US-09-804-682-33/c  
; Sequence 33, Application US/09804682  
; Patent No. US20020106765A1

## GENERAL INFORMATION:

;; APPLICANT: Kinders, Robert

```
; APPLICANT: Corey, Michael J.
; TITLE OF INVENTION: PAL-18 POLYPEPTIDES, NUCLEIC ACIDS
; TITLE OF INVENTION: ENCODING THE SAME AND METHODS FOR SCREENING FOR OR
; TITLE OF INVENTION: MODULATING THE SAME
; FILE REFERENCE: 130001.406
; CURRENT APPLICATION NUMBER: US/09/804,682
; CURRENT FILING DATE: 2001-03-12
; NUMBER OF SEQ ID NOS: 174
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 33
; LENGTH: 1065
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 14, 21, 27, 33, 36, 42, 72, 101, 103, 207, 208, 221, 223,
; LOCATION: 227, 235, 240, 242, 243, 247, 248, 259, 263, 269, 273, 278,
; LOCATION: 296, 321, 322, 324, 330, 332, 335, 336, 340, 367, 371, 385,
; LOCATION: 390, 393, 399, 401, 407, 415, 421, 447, 454, 471, 475
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 479, 494, 501, 508, 511, 513, 525, 539, 540, 579, 596, 605,
; LOCATION: 623, 624, 633, 653, 654, 656, 658, 664, 670, 679, 711, 713,
; LOCATION: 721, 729, 735, 743, 747, 748, 754, 776, 778, 779, 780, 783,
; LOCATION: 798, 802, 808, 810, 813, 814, 820, 822, 824, 825, 838
; OTHER INFORMATION: n = A,T,C or G
; NAME/KEY: misc_feature
; LOCATION: 842, 847, 888, 900, 909, 910, 916, 926, 927, 943, 948, 962,
; LOCATION: 973, 1002, 1005, 1028, 1029, 1034, 1057, 1065
; OTHER INFORMATION: n = A,T,C or G
; US-09-804-682-33

Query Match 3.8%; Score 38; DB 10; Length 1065;
Best Local Similarity 47.2%; Pred. No. 0.04;
Matches 83; Conservative 0; Mismatches 93; Indels 0; Gaps 0;

QY 224 GGCAGTGTGTGGTGGAGTGAAGACAGACGCGGTGGGAGGTCAGGAGAGAATCTCT 283
DB 742 GGGGGGNGGGGNGGGAAGNGGGGAATNGNGAGGGGAGGGGGGGGAGAGGAGGA 683

QY 284 GGCTGGGGATGTGTGGGATCACTGTCCATTGTGCAGGCTGGTCTTGGGCGAGG 343
DB 682 GGGTGGGAAGNGGGGNGAGGNGGNNNAAGGGGGGGGGGGTNGGGGGGGGNN 623

QY 344 AAGGGATGGGGGCCATAGCAGTGTGGTGCAGCGAGCTGGCTGGGAAGTGGTG 399
DB 622 GAGGGGGGGGGGGGNGGGGGGNGGAGGGGAGGGGGGNGGGGAGGGGGG 567

RESULT 8
US-09-970-921-5/c
; Sequence 5, Application US/09970921
; Patent No. US20020133845A1
; GENERAL INFORMATION:
; APPLICANT: Frank Michiels et al.
; TITLE OF INVENTION: Improved barstar Gene
; FILE REFERENCE: 2428-0108P
; CURRENT APPLICATION NUMBER: US/09/970,921
; CURRENT FILING DATE: 2001-10-05
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 4032
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: other nucleic
; OTHER INFORMATION: acid, "plasmid pmv71"
; NAME/KEY: misc_feature
; LOCATION: (1999)..(3400)
; OTHER INFORMATION: label = PRACL, "promoter region of rice actin gene
; OTHER INFORMATION: - contains an intron in the leader"
; NAME/KEY: misc_feature
```

```
; LOCATION: (3401)..(3676)
; OTHER INFORMATION: label = barstar, "barstar DNA"
; NAME/KEY: misc_feature
; LOCATION: (3677)..(4003)
; OTHER INFORMATION: label = 3'nos, "region containing 3' untranslated
; OTHER INFORMATION: end of nopaline synthase gene of Agrobacterium
; OTHER INFORMATION: T-DNA
; NAME/KEY: misc_feature
; LOCATION: (3399)..(3404)
; OTHER INFORMATION: label = NcoI, "NcoI recognition site"
; NAME/KEY: misc_feature
; LOCATION: (4016)..(4021)
; OTHER INFORMATION: label = KpnI, "KpnI recognition site"
; US-09-970-921-5

Query Match 3.8%; Score 38; DB 10; Length 4032;
Best Local Similarity 50.4%; Pred. No. 0.063;
Matches 118; Conservative 0; Mismatches 115; Indels 1; Gaps 1;

QY 120 CAGGCCCCCAACCCGCCGCCAGACACTGTCTCAGGCCCTCTCAGCCTGTCTGATTT 179
DB 2724 CGGGCCCCCAGACCCGGACCCACCTGCTGTTTTTCTTTTCTTTTCTTTTCTTTT 2665

QY 180 GCTTCTGTGCTCGGGGAGAAATGAGTGGGAGAAAACACAGCCAGGCGAGTTGGTGG 239
DB 2664 TTTTCTTTTGGTGGGAGACGTGCGGTGCGGACAACTCACGGTGTAGTGGGGGG 2605

QY 240 AGTGAAGACGACGCGGTGGGAGGTCAGGAGAGAACTCTCTGGCTGGGATGGTGT 299
DB 2604 TGTGGAGACTATGTCTCAGTTGGTGGAGTGGGTTGGTGGTGGTGGTGGTGGCT 2545

QY 300 GGGCATCAACTCTCCCATTTGCT-GCAGGCTGGTCTTTGGGCGAGGGAGGGATG 352
DB 2544 GGGCTTGTATGATGATCGTGGATAGCACATTGGCTTAGGAACTTTAGGGGTG 2491

RESULT 9
US-09-904-968A-1/c
; Sequence 1, Application US/09904968A
; Publication No. US20030008288A1
; GENERAL INFORMATION:
; APPLICANT: THE JOHNS HOPKINS UNIVERSITY SCHOOL OF MEDICINE
; APPLICANT: GERMINO, Gregory
; APPLICANT: WATNICK, Terry
; APPLICANT: PHAKDEKITCHAROEN, Bunyong
; TITLE OF INVENTION: DETECTION AND TREATMENT OF POLYCYSTIC KIDNEY DISEASE
; FILE REFERENCE: JHU1680-2
; CURRENT APPLICATION NUMBER: US/09/904,968A
; CURRENT FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/283,691
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: US 60/218,261
; PRIOR FILING DATE: 2000-07-13
; NUMBER OF SEQ ID NOS: 113
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 1
; LENGTH: 53522
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-904-968A-1

Query Match 3.7%; Score 37.4; DB 9; Length 53522;
Best Local Similarity 52.2%; Pred. No. 0.24;
Matches 83; Conservative 0; Mismatches 76; Indels 0; Gaps 0;

QY 199 AATGAGTGGGAGAAAACAGCCAGCGGAGTTGGTGTGGAGTGAGAGACAGCGCGG 258
DB 35656 AAAGGAAAAGGGGAGAGAGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 35597

QY 259 TGGGAGGTCAGGAGAGAAATCTGTGGCTGGGATGGTGGGATCACTGTCCTT 318
DB 35596 AGGGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 35537
```







GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:06:43 ; Search time 1290.57 Seconds  
(without alignments)  
12561.622 Million cell updates/sec

Title: US-09-691-220-3-copy\_5000\_6000

Perfect score: 1001

Sequence: 1 tgatctccctcccccacc.....tgggtaccctctctggtg 1001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*\*

1: em\_estba:\*\*

2: em\_esthum:\*\*

3: em\_estin:\*\*

4: em\_estmu:\*\*

5: em\_estov:\*\*

6: em\_estpl:\*\*

7: em\_estro:\*\*

8: em\_htc:\*\*

9: gb\_est1:\*\*

10: gb\_est2:\*\*

11: gb\_htc:\*\*

12: gb\_est3:\*\*

13: gb\_est4:\*\*

14: gb\_est5:\*\*

15: em\_estfun:\*\*

16: em\_estom:\*\*

17: gb\_gss:\*\*

18: em\_gss\_hum:\*\*

19: em\_gss\_inv:\*\*

20: em\_gss\_pln:\*\*

21: em\_gss\_vrt:\*\*

22: em\_gss\_fun:\*\*

23: em\_gss\_mam:\*\*

24: em\_gss\_mus:\*\*

25: em\_gss\_other:\*\*

26: em\_gss\_pro:\*\*

27: em\_gss\_rod:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
c 1	74	7.4	1101	17	CNS00397
c 2	50	5.0	974	17	BH157536
c 3	49.8	5.0	620	17	AG100622
c 4	49.4	4.9	1046	14	BQ682356
c 5	48.4	4.8	1398	13	BM479595
c 6	47.8	4.8	817	17	CNS041N8
c 1	74	7.4	1101	17	CNS00397
c 2	50	5.0	974	17	BH157536
c 3	49.8	5.0	620	17	AG100622
c 4	49.4	4.9	1046	14	BQ682356
c 5	48.4	4.8	1398	13	BM479595
c 6	47.8	4.8	817	17	CNS041N8

7	47.8	4.8	910	17	BH162948
c 8	47.8	4.8	974	17	BH157536
c 9	47.8	4.8	997	17	CNS006DN
10	46.4	4.6	384	17	CNS035G7
11	46.4	4.6	1101	17	CNS017ZR
c 12	46.2	4.6	936	17	CNS01608
c 13	46	4.6	884	17	CNS00600
c 14	45.8	4.6	951	9	AL543262
c 15	45.6	4.6	758	17	AG080593
c 16	45.6	4.6	895	17	AZ667655
c 17	45.6	4.6	1101	17	CNS0020T
c 18	45.4	4.5	1249	12	BE873655
c 19	45.4	4.5	1417	14	BM810236
c 20	45.2	4.5	897	17	AZ186654
c 21	45.2	4.5	1040	17	CNS01671
c 22	45	4.5	544	17	CNS015XA
c 23	45	4.5	706	17	AG062296
c 24	45	4.5	724	17	AZ197686
c 25	45	4.5	730	17	AG076255
c 26	45	4.5	1113	17	AG030979
c 27	44.8	4.5	324	17	CNS001CM
c 28	44.8	4.5	894	17	AG081742
c 29	44.8	4.5	925	17	CNS0091P
c 30	44.8	4.5	962	17	AG152777
c 31	44.6	4.5	506	9	AL549672
c 32	44.6	4.5	857	17	BH157466
c 33	44.4	4.4	1047	17	CNS006KX
c 34	44.4	4.4	1282	14	BQ65718
c 35	44.2	4.4	556	17	CNS046C5
c 36	44.2	4.4	777	17	AG104642
c 37	44.2	4.4	923	9	AL543393
c 38	44.2	4.4	925	17	CNS0091P
c 39	44.2	4.4	1041	17	CNS00LFU
c 40	44.2	4.4	1201	9	AL581589
c 41	44	4.4	461	17	BH750315
c 42	44	4.4	694	17	AG100694
c 43	44	4.4	895	17	CNS06LLW
c 44	44	4.4	996	17	CNS04CW3
c 45	44	4.4	1101	17	CNS00K66

## ALIGNMENTS

RESULT 1	CNS00397	1101 bp	DNA	linear	GSS 03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence T7 end of BAC #				
DEFINITION	BACR08A10 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.				
ACCESSION	AL063912				
VERSION	AL063912.1	GI:4941769			
KEYWORDS	GSS.				
SOURCE	Drosophila melanogaster.				
ORGANISM	Drosophila melanogaster				
	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.				
REFERENCE	Genoscope.				
AUTHORS	Direct Submission				
TITLE	Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :				
JOURNAL	BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr)				
COMMENT	<p>Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <a href="http://www.fruitfly.org">http://www.fruitfly.org</a> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial</p>				















```

Db      150 GGGGTGKRRARGGGGRKKRRTKGK 122          |||:::~::~|||

RESULT 13
LOCUS   CNS006UO                                884 bp    DNA        linear     GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC #BACRI4N21 of RpCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL065923
VERSION   AL065923.1 GI:4944891
KEYWORDS  GSS.
SOURCE    Drosophila melanogaster.
ORGANISM Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephedroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 884)
AUTHORS
TITLE     Direct Submission
JOURNAL
COMMENT   Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr ; Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org/The_BDGP_Drosophila_melanogaster_BAC_library.html prepared by Kazutoyo Osoegawa and Aaron Mammosier in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RpCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp; the same strain used for the BDGP's pl and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES             source
BASE COUNT           230 a   62 c   139 g   124 t   329 others
ORIGIN                Location/Qualifiers
                    1..884
                        /organism="Drosophila melanogaster"
                        /db_xref="taxon:7227"
                        /clone="BACRI4N21"
                        /clone_lib="RpCI-98"
                        /notes="end : T7"

Query Match         4.6% ; Score 46 ; DB 17 ; Length 884 ;
Best Local Similarity 19.5% ; Pred. No. 0.18 ;
Matches            44 ; Conservative 93 ; Mismatches 89 ; Indels 0 ; Gaps 0 ;

QY 174 TGATTTCCTTGCTGGCGCTGGGAGATGAGTGGGAGAAAACCAGGCCAGGCAGTTGG 233
       || | | :::: || ||::: || | | | | | | | | | | | | | | | | 
Db 623 TGKTGTGTGTTGGGGGGGGGKXKKRGSGTKGSGGGGKKGSTKTGTGDKDTGGKKKTKT 682
       QY 234 TTGTGGAGTGAAGACAGACGGCGGTGGGAGGTGCAGAGAGAATCTGTGGCTGGGA 293
       :::::: || | | | | | | | | | | | | | | | | | | | | | | 
Db 683 KXXXXXXGTXGKNGXTFKXSOTYKKKGXTGTGTGVTGTTTTKKXGKRGKRKGKXGX 742
       QY 294 TGGTGTGGCATCAACTGTCCTCCCATTCGTGCAGGCTGGTCTTGGGGCAGGAAGGGATGG 353
       ||:::|| | | | | | | | | | | | | | | | | | | | | | | 
Db 743 KGKKRKRGKKKKKKKKKKKKKTKTGTKGKGTGKKKKKKKTKKKKKKKKKKKKK 802
       QY 354 GGCGCCATACAGTGTGTCGCCAGCGTGGCTGGGAAGTGGTG 399
       ||:::|| | | | | | | | | | | | | | | | | | | | | | | 
Db 803 KKXXXXXXXXXXXXXXXXXXXKTKKKKKKKCKKKKKKKKKKKKK 848

RESULT 14
AL543262/C
LOCUS    AL543262                               951 bp    mRNA        linear     EST 16-FEB-2001
```



Search completed: March 30, 2003, 13:48:16  
Job time : 1300.57 secs



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:00:38 ; Search time 2122.53 Seconds  
(without alignments)  
13725.086 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_10000\_11000  
Perfect score: 1001  
Sequence: 1 cccaggagactgcagctgg.....aaggctgcctgcctggacat 1001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_man.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
C 1	1001	100.0	138999	9	AC090426	AC090426 Homo sapi
2	1001	100.0	158766	2	AC015851	AC015851 Homo sapi
3	1001	100.0	166368	2	AC018629	AC018629 Homo sapi
C 4	1001	100.0	173441	2	AC080112	AC080112 Homo sapi
C 5	1001	100.0	188574	2	AC131063	AC131063 Homo sapi
6	1001	100.0	190309	2	AC126392	AC126392 Homo sapi
7	997.8	99.7	1233	9	HSRARA1A5	AF088892 Homo sapi
8	255.6	25.5	231248	2	AL591067	AF088892 Homo sapi
C 9	204.4	20.4	162393	2	AC111746	AC111746 Rattus no
10	174	17.4	704	6	AR052138	AR052138 Sequence
11	174	17.4	1284	9	S50916	S50916 PML-RAR fus
12	174	17.4	1572	12	AF242868	AF242868 Synthetic
13	174	17.4	1920	9	HSRAR	X06538 Human mRNA
14	174	17.4	1944	9	HSU41743	U41743 Human nucle
15	174	17.4	2008	9	AK098172	AK098172 Homo sapi
16	174	17.4	2073	9	HSU41742	U41742 Human nucle
17	174	17.4	2085	12	AF242867	AF242867 Synthetic
18	174	17.4	2441	9	BC008727	BC008727 Homo sapi
19	174	17.4	2658	6	AR052137	AR052137 Sequence
20	174	17.4	2907	6	AX333042	AX333042 Sequence
21	174	17.4	2907	9	HSRRA	X06614 Human mRNA
22	174	17.4	2928	6	AR061122	AR061122 Sequence
23	174	17.4	2940	6	AR052136	AR052136 Sequence
24	174	17.4	2940	6	I09348	I09348 Sequence 1
25	174	17.4	3036	6	AR061121	AR061121 Sequence
26	174	17.4	3036	6	I96215	I96215 Sequence 52
27	174	17.4	3036	9	HUMPLRAR	M73779 Human PML-R
28	172.4	17.2	2908	6	I08117	I08117 Sequence 1
29	167.6	16.7	1474	10	MSRARA	Y10094 M.spretus m
30	167.6	16.7	1759	10	MURARAP	M60909 Mouse retin
31	167.6	16.7	1926	10	MNRARA	X57528 M.musculus
32	167.6	16.7	2061	10	S56656	S56656 retinolic ac
33	166	16.6	908	10	RNAJ2940	AJ002940 Rattus no
34	166	16.6	924	10	RNAJ2941	AJ002941 Rattus no
35	166	16.6	2130	10	RNU15211	U15211 Rattus norv
36	162.8	16.3	2748	10	AY046943	AY046943 Mesocric
37	161.4	16.1	2300	9	HSR805408	AL834159 Homo sapi
C 38	156	15.6	301	6	AX098489	AX098489 Sequence
39	151.8	15.2	392	4	SSU82260	U82260 Sus scrofa
40	143.2	14.3	1494	5	GGRARA21	X78335 G.gallus re
41	143.2	14.3	1552	5	GGRARA1	X78372 G.gallus RA
42	138.8	13.9	860	5	AB073783	AB073783 Paralicht
43	137.8	13.8	1377	5	NVRARAG	X17585 Newt mRNA f
44	137.8	13.8	1697	5	NVRARA2M	Z14254 N.viridesc
C 45	132.2	13.2	110000	2	AC021072_2	Continuation (3 of

ALIGNMENTS

RESULT 1  
AC090426/c  
LOCUS  
DEFINITION Homo sapiens chromosome 17 clone 205m17 map 17q21.1, complete  
ACCESSION AC090426  
VERSION AC090426.1 GI:12965315  
KEYWORDS HTG.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 138999)  
AUTHORS Gu,B., Xiong,H., Zhou,Y., Chen,B., Lu,L., Zhong,M., Yin,H.,  
Huang,W., Ren,S., Chen,S.F., Chen,Z. and Fu,G.

TITLE	Variant-type PML-RAR(alpha) fusion transcript in acute promyelocytic leukemia: use of a cryptic coding sequence from intron 2 of the RAR(alpha) gene and identification of a new clinical subtype resistant to retinoic acid therapy
JOURNAL MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 99 (11), 7640-7645 (2002)
PUBMED	12028997
REFERENCE	12032336
AUTHORS	2 (bases 1 to 138999) Gu,B., Xiong,H., Zhou,Y., Chen,B., Lu,L., Zhong,M., Yin,H., Huang,W., Ren,S., Chen,S.F., Chen,Z. and Fu,G.
TITLE	Direct Submission
JOURNAL	Submitted (20-FEB-2001) Chinese National Human Genome Center at Shanghai, Shanghai, Shanghai 201203, P.R.China
COMMENT	-----Genome Center----- Center:Chinese National Human Genome Center at Shanghai Center code:CHGC Website: http://www.chgc.sh.cn Contact: fufang@chgc.sh.cn
FEATURES	Location/Qualifiers
source	1..138999
	/organism="Homo sapiens"
	/db_xref="taxon:9606"
	/chromosome="17"
	/map="17q21.1"
	/clone="205m17"
repeat_region	387..483
	/rpt_family="SINE/MIR"
	/rpt_unit="MIR"
repeat_region	725..1032
	/rpt_family="SINE/Alu"
	/rpt_unit="AluJb"
repeat_region	1033..1198
	/rpt_family="SINE/Alu"
repeat_region	1217..1514
	/rpt_family="SINE/Alu"
	/rpt_unit="AluSp"
repeat_region	complement(1534..1754)
	/rpt_family="SINE/MIR"
	/rpt_unit="MIR"
repeat_region	2291..2366
	/rpt_family="SINE/MIR"
	/rpt_unit="MIR"
repeat_region	complement(2607..2665)
	/rpt_family="SINE/MIR"
	/rpt_unit="MIR"
repeat_region	3631..3905
	/rpt_family="LINE/L1"
	/rpt_unit="L1ME2"
repeat_region	3977..4029
	/rpt_family="LINE/L1"
	/rpt_unit="L1ME2"
repeat_region	4030..4213
	/rpt_family="LTR/MaLR"
	/rpt_unit="MSTD"
repeat_region	4214..4509
	/rpt_family="SINE/Alu"
	/rpt_unit="AluJb"
repeat_region	4514..4811
	/rpt_family="SINE/Alu"
	/rpt_unit="AluSx"
repeat_region	4812..5018
	/rpt_family="LTR/MaLR"
	/rpt_unit="MSTD"
repeat_region	5019..5102
	/rpt_family="LINE/L1"
	/rpt_unit="L1ME2"
repeat_region	5236..5443
	/rpt_family="SINE/MIR"
	/rpt_unit="MIR"
repeat_region	5467..5699
	/rpt_family="SINE/MIR"
	/rpt_unit="MIR"
repeat_region	6547..6708
	/rpt_family="SINE/MIR"
	/rpt_unit="MIR"
	/rpt_unit="MIR"
	6827..6917
	/rpt_family="Low_complexity"
	/rpt_unit="GA-rich"
repeat_region	7733..7779
	/rpt_family="SINE/MIR"
	/rpt_unit="MIR"
repeat_region	8268..8411
	/rpt_family="Simple_repeat"
	/rpt_unit="(TGGG)n
repeat_region	9224..9532
	/rpt_family="SINE/Alu"
	/rpt_unit="AluY"
repeat_region	9546..9639
	/rpt_family="Low_complexity"
	/rpt_unit="CT-rich"
repeat_region	10025..10426
	/rpt_family="LTR/Retroviral"
	/rpt_unit="MLT2A"
repeat_region	10433..10471
	/rpt_family="Simple_repeat"
	/rpt_unit="(TA)n
repeat_region	10472..10527
	/rpt_family="LTR/Retroviral"
	/rpt_unit="MLT2A"
repeat_region	complement(10675..10962)
	/rpt_family="SINE/Alu"
	/rpt_unit="AluY"
repeat_region	11013..11321
	/rpt_family="SINE/Alu"
	/rpt_unit="AluY"
repeat_region	complement(11840..12192)
	/rpt_family="LTR/MaLR"
	/rpt_unit="MSTD"
repeat_region	12729..12835
	/rpt_family="SINE/Alu"
	/rpt_unit="FLAM_A"
repeat_region	complement(12841..12891)
	/rpt_family="LINE/L2"
	/rpt_unit="L2"
repeat_region	complement(14228..14548)
	/rpt_family="SINE/Alu"
	/rpt_unit="AluY"
repeat_region	complement(14618..14701)
	/rpt_family="SINE/MIR"
	/rpt_unit="MIR"
repeat_region	14953..15255
	/rpt_family="SINE/Alu"
	/rpt_unit="AluJb"
repeat_region	15311..15473
	/rpt_family="Simple_repeat"
	/rpt_unit="(TA)n
repeat_region	15669..15981
	/rpt_family="SINE/Alu"
	/rpt_unit="AluSx"
repeat_region	16316..16353
	/rpt_family="Simple_repeat"
	/rpt_unit="(TAAA)n
repeat_region	complement(16863..16910)
	/rpt_family="SINE/MIR"
	/rpt_unit="MIR"
repeat_region	17026..17324
	/rpt_family="SINE/Alu"
	/rpt_unit="AluSg"
repeat_region	17337..17618
	/rpt_family="SINE/Alu"
	/rpt_unit="AluSx"
repeat_region	17619..18024
	/rpt_family="LTR"
	/rpt_unit="LPR54"
repeat_region	complement(18026..18086)
	/rpt_family="SINE/MIR"
	/rpt_unit="MIR"







```

QY 481 GCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGCAAAAGCGCACAGGAACCTTCCCTCG 540
D 7025 GCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGCAAAAGCGCACAGGAACCTTCCCTCG 7084
QY 541 CCTCTGCCAGCTGGGCAAAATACACTACGATGCTTTTCCCGGCTTCGAGGTGGGAT 600
D 7085 CCTCTGCCAGCTGGGCAAAATACACTACGATGCTTTTCCCGGCTTCGAGGTGGGAT 7144
QY 601 TTGCCAGGCGCACAGGCGCAGGATGGGCGCTCTCAGGACACCGCTTCTTGTGCGAGCA 660
D 7145 TTGCCAGGCGCACAGGCGCAGGATGGGCGCTCTCAGGACACCGCTTCTTGTGCGAGCA 7204
QY 661 AGATCTCTGGTCTTCCCTTCCCTCTCTTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 720
D 7205 AGATCTCTGGTCTTCCCTTCCCTCTCTTCTCCCTCTCTCTCTCTCTCTCTCTCTCTCT 7264
QY 721 AGTCCAGGAAGTGAAGGCTGGGTAGAGGCGAGCGCTGTGGGGCTGGAGCCAGGCTGA 780
D 7265 AGTCCAGGAAGTGAAGGCTGGGTAGAGGCGAGCGCTGTGGGGCTGGAGCCAGGCTGA 7324
QY 781 GAGGGGTGCATGGAGAAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
D 7325 GAGGGGTGCATGGAGAAGGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7384
QY 841 GAACAAGTCTCTCTCTGACATGACCTCTGGGACAGTTCAGTGAAGTCTCCACCAAG 900
D 7385 GAACAAGTCTCTCTCTGACATGACCTCTGGGACAGTTCAGTGAAGTCTCCACCAAG 7444
QY 901 TGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCGCGCTTCACCAACCTCACCAC 960
D 7445 TGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCGCGCTTCACCAACCTCACCAC 7504
QY 961 GCGGACAGATCACCTCTCTCAAGGCTGCGCGCTTCGACAT 1001
D 7505 GCGGACAGATCACCTCTCTCAAGGCTGCGCGCTTCGACAT 7545

RESULT 4
AC080112/c
LOCUS
DEFINITION
IN PROGRESS ***, 3 unordered pieces.
AC080112
AC080112.6 GI:2218574
HTG: HTGS_PHASE1; HTGS_FULLTOP; HTGS_ACTIVEFIN.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 173441)
Birren, B., Nussbaum, C. and Lander, E.
Homo sapiens chromosome 17, clone CTD-2267D19 map 17, *** SEQUENCING.
Unpublished
2 (bases 1 to 173441)
Birren, B., Linton, J., Nussbaum, C., Lander, E., Abraham, H., Allen, N.,
Anderson, S., Barna, N., Bastien, V., Bieda, F., Boguslavskiy, L.,
Boukhgalter, B., Brown, A., Burkett, G., Campolano, A., Castle, A.,
Chapel, Y., Collangelo, M., Collins, S., Collymore, A., Cooke, P.,
DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S., Ferreira, P.,
Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M.,
Graham, L., Grand-pierre, N., Hagos, B., Heaford, A., Horton, L.,
Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Lacombe, K.,
Lamazares, R., Landers, T., Lehoczy, J., Levine, R., Lieu, C., Liu, G.,
Macdonald, P., Marquis, N., McCarthy, M., McEwan, P., McKernan, K.,
McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,
Morrow, J., Murphy, T., Naylor, J., Norman, C.H., O'Connor, T.,
O'Donnell, P., O'Neill, D., Pollara, V., Raymond, C., Rieback, M., Riley, R.,
Pierre, N., Pisani, C., Pollara, V., Raymond, C., Schauer, S., Severy, P.,
Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P.,
Sougnuez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,
Tirrell, A., Travers, M., Trigilio, J., Vassiliev, H., Viel, R., Vo, A.,
Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,

```

```

TITLE
JOURNAL
REFERENCE
AUTHORS
Zimmer, A. and Zody, M.
Direct Submission
Submitted (24-SEP-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 173441)
Birren, B., Nussbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,
Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cooke, P., Cooke, P., DeArelano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Ginde, S., Grand-pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., Maclean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schupback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Direct Submission
Submitted (14-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 14, 2002 this sequence version replaced 91:22123315.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
----- Project Information
Center project name: l11030
Center clone name: 2267_D_19
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 3 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 2128: contig of 2128 bp in length
* 2129 2228: gap of 100 bp
* 2229 157931: contig of 157903 bp in length
* 157932 158031: gap of 100 bp
* 158032 173441: contig of 15410 bp in length.
* Location/Qualifiers
* 1. 173441
* /organism="Homo sapiens"
* /db_xref="taxon:9606"
* /chromosome="17"
* /map="17"
* /clone="CTD-2267D19"
* /clone.lib="CITD1 Human BAC"
BASE COUNT 46772 a 43344 c 40918 g 42114 t 293 others
ORIGIN
Query Match 100.0%; Score 1001; DB 2; Length 173441;
Best Local Similarity 100.0%; pred. No. 5.3e-206;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1 CCCAGGAGACTGCAGCTGGAGGCGCTGGGTGAGTGGAGCGGAGAGACCTTCCTCG 60
D 72235 CCCAGGAGACTGCAGCTGGAGGCGCTGGGTGAGTGGAGCGGAGAGACCTTCCTCG 72176
Qy 61 GGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 120
D 120 GGAAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG

```





```
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-1029L16"
/clone_lib="RPC1-11 Human Male BAC"
1. .2072
misc_feature
    /note="assembly_fragment"
    clone_end:SP6
    vector_side:left
2173. .19348
misc_feature
    /note="assembly_fragment"
    19449. .34798
misc_feature
    /note="assembly_fragment"
    34899. .74660
misc_feature
    /note="assembly_fragment"
    74761. .151113
misc_feature
    /note="assembly_fragment"
    151214. .188574
misc_feature
    /note="assembly_fragment"
    clone_end:T7
    vector_side:right"
BASE COUNT 48621 a 48660 c 47029 g 43760 t 504 others
ORIGIN
Query Match 100.0%; Score 1001; DB 2; Length 188574;
Best Local Similarity 100.0%; Pred. No. 5.2e-206;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 CCCAGGAGACTGCAGCTGGAGGGCTGGGTGAGTGGAGGGGAGGAGGACCTTCCTGG 60
Db 155844 CCCAGGAGACTGCAGCTGGAGGGCTGGGTGAGTGGAGGGGAGGAGGACCTTCCTGG 155785
QY 61 GGAAGAGGAGGAGACCTAGAGGGGACCGCTGCGCTGGAGTGTGAGCTGGAGTGA 120
Db 155784 GGAAGAGGAGGAGACCTAGAGGGGACCGCTGCGCTGGAGTGTGAGCTGGAGTGA 155725
QY 121 CGCTGGGGATACATCGCTGCGCTGATGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 180
Db 155724 CGCTGGGGATACATCGCTGCGCTGATGGGTGGGTGGGTGGGTGGGTGGGTGGGTGG 155665
QY 181 AGCTGTCTCATGGGGCTTCGGGGCAGAACTTGATGTGTGGGTGGGTGGGTGGGTGGGTGG 240
Db 155664 AGCTGTCTCATGGGGCTTCGGGGCAGAACTTGATGTGTGGGTGGGTGGGTGGGTGGGTGG 155605
QY 241 GCTGGAGTGGTGCATGCTTCCTGCTGCGGTGAACTGCTGTGTGGTGGTGGTGGTGGTGGTGG 300
Db 155604 GCTGGAGTGGTGGCAATGCTTCCTGCTGCGGTGAACTGCTGTGTGGTGGTGGTGGTGGTGG 155545
QY 301 AAGCCTGGGTGACCTCTCAGCAGCTGCGAGCTCTCTGTCAAGCTGGGTGGGTGGGTGGGTGG 360
Db 155544 AAGCCTGGGTGACCTCTCAGCAGCTGCGAGCTCTCTGTCAAGCTGGGTGGGTGGGTGGGTGG 155485
QY 361 CTTGAGCAGCTGCAGTGGCTTAAACCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
Db 155484 CTTGAGCAGCTGCAGTGGCTTAAACCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 155425
QY 421 CCGAAACAAGAGAGAGAGAGTGGCCCAAGCCGAGTGTCTGAGAGCTACAGCTGAC 480
Db 155424 CCGAAACAAGAGAGAGAGTGGCCCAAGCCGAGTGTCTGAGAGCTACAGCTGAC 155365
QY 481 CGCGAGGTGGGGAGCTCTATTGAGAAGTGGCAAGAGCGCACAGGAAACCTTCCTCCCTGG 540
Db 155364 CGCGAGGTGGGGAGCTCTATTGAGAAGTGGCAAGAGCGCACAGGAAACCTTCCTCCCTGG 155305
QY 541 CTTCTGCCAGCTGGGCAATACACTACGGTATGCTTTTCCCGCCCTGCGAGGTGGAT 600
Db 155304 CTTCTGCCAGCTGGGCAATACACTACGGTATGCTTTTCCCGCCCTGCGAGGTGGAT 155245
QY 601 TTGCCAGGCGCCACAGGCGGAGGATGGCCCTCTCAGGACACCCCTCTCTTGTGCGCAGGCA 660
Db 155244 TTGCCAGGCGCCACAGGCGGAGGATGGCCCTCTCAGGACACCCCTCTCTTGTGCGCAGGCA 155185
QY 661 AGATCTCTGCTGCTTCCCTTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
```

```
Db 155184 AGATCTCTGCTGCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCCCTTCC 155125
QY 721 AGCTCCCAAGAGTGAAGGCTGGGTAGAGGGCAGGCTGTGGGGGCTGGAGCCAGGCTGA 780
Db 155124 AGCTCCCAAGAGTGAAGGCTGGGTAGAGGGCAGGCTGTGGGGGCTGGAGCCAGGCTGA 155065
QY 781 GAAGGGTGGCCATGGAGAGAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 840
Db 155064 GAAGGGTGGCCATGGAGAGAGAGGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 155005
QY 841 GAACAACGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900
Db 155004 GAACAACGTGTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 154945
QY 901 TGCATCATTAAGACTGTGTGGAGTTCGCCAAGCAGCTGCCGGCTTACCCACCTCCACCATC 960
Db 154944 TGCATCATTAAGACTGTGTGGAGTTCGCCAAGCAGCTGCCGGCTTACCCACCTCCACCATC 154885
QY 961 GCCGACCAGATCACCTCTCTCAAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1001
Db 154884 GCCGACCAGATCACCTCTCTCAAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 154844

RESULT 6
AC126392 190309 bp DNA linear HTG 06-AUG-2002
LOCUS Homo sapiens chromosome 17 clone RP11-1029F16 map 17, WORKING DRAFT
DEFINITION AC126392
ACCESSION AC126392
VERSION AC126392.2 GI:22123095
KEYWORDS HTG; HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 190309)
Birren,B., Nusbaum,C. and Lander,E.
Homo sapiens chromosome 17, clone RP11-1029p16
Unpublished
2 (bases 1 to 190309)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schupback,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,A. and Zody,M.
Direct Submission
Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 190309)
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavskiy,L., Boukhalter,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzgerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kellis,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
```



Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission  
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 6, 2002 this sequence version replaced gi:21699357.  
All repeats were identified using RepeatMasker:  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

# TITLE JOURNAL

## COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu

## Project Information

Center project name: L27553

Center clone name: 1029\_F16

----- Summary Statistics

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 187521 bases at least Q40

Consensus quality: 188193 bases at least Q30

Consensus quality: 188537 bases at least Q20

Insert size: 179000; agarose-fp

Insert size: 189309; sum-of-contigs

Quality coverage: 15.5 in Q20 bases; agarose-fp

Quality coverage: 14.6 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 11 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.

\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 30513: contig of 30513 bp in length  
\* 30514 30613: gap of 100 bp  
\* 30614 31764: contig of 1151 bp in length  
\* 31765 31864: gap of 100 bp  
\* 31865 33693: contig of 1829 bp in length  
\* 33694 33793: gap of 100 bp  
\* 33794 35405: contig of 1612 bp in length  
\* 35406 35505: gap of 100 bp  
\* 35506 37691: contig of 2186 bp in length  
\* 37692 37791: gap of 100 bp  
\* 37792 51370: contig of 13579 bp in length  
\* 51371 51470: gap of 100 bp  
\* 51471 66014: contig of 14544 bp in length  
\* 66015 66114: gap of 100 bp  
\* 66115 108183: contig of 42069 bp in length  
\* 108184 108283: gap of 100 bp  
\* 108284 130510: contig of 22227 bp in length  
\* 130511 130610: gap of 100 bp  
\* 130611 166740: contig of 36130 bp in length  
\* 166741 166840: gap of 100 bp  
\* 166841 190309: contig of 23469 bp in length.

## FEATURES

### source

1. 190309  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/map="17"  
/chromosomes="17"  
/clone="RP11-1029F16"  
/clone\_lib="RPC1-11 Human Male BAC"

misc\_feature 1. 30513  
clone\_end:SP6  
vector\_side:left  
30614. 31764  
/note="assembly\_fragment"  
31865. 33693  
/note="assembly\_fragment"  
33794. 35405  
/note="assembly\_fragment"  
35506. 37691  
/note="assembly\_fragment"  
37792. 51370  
/note="assembly\_fragment"  
51471. 66014  
/note="assembly\_fragment"  
66115. 108183  
/note="assembly\_fragment"  
108284. 130510  
/note="assembly\_fragment"  
130611. 166740  
/note="assembly\_fragment"  
166841. 190309  
/note="assembly\_fragment"  
clone\_end:T7  
vector\_side:right"

BASE COUNT 48207 a 48484 c 47543 g 45018 t 1057 others  
ORIGIN

Query Match 100.0%; Score 1001; DB 2; Length 190309;  
Best Local Similarity 100.0%; Pred. No. 5.2e-206;  
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 CCCAGGGAGACTGCAGCTGGGAGGCTGGGTGAGTGGAGGCGGAGAGAGACCTTCCTGG 60  
Db 46671 CCCAGGAGACTGCAGCTGGGAGGCTGGGTGAGTGGAGGCGGAGAGAGACCTTCCTGG 46730  
Qy 61 GGAAAGAGGAGGAGAGACCTTAGAGGGACCCGCTGCGCTGGAGTGTGAGCTGGAGTAGA 120  
Db 46731 GGAAAGAGGAGGAGAGACCTTAGAGGGACCCGCTGCGCTGGAGTGTGAGCTGGAGTAGA 46790  
Qy 121 CGCCTGGGGATAGCATCGCGCTGCCTATAGGGTGGGGTGGGGTGTGTGCAGGGCCAC 180  
Db 46791 CGCCTGGGGATAGCATCGCGCTGCCTATAGGGTGGGGTGGGGTGTGTGCAGGGCCAC 46850  
Qy 181 AGCTGTGCTCATGGGGCTTCCTGGGCGACAACTTGATGTGTGGTGGGTGGGCATGGAGG 240  
Db 46851 AGCTGTGCTCATGGGGCTTCCTGGGCGACAACTTGATGTGTGGTGGGTGGGCATGGAGG 46910  
Qy 241 GCTGGAGTGCCTGGCAATGCCCTTGCCTCCCGTGAACGCGTGTCTGTGCGCGTGTCTTAC 300  
Db 46911 GCTGGAGTGCCTGGCAATGCCCTTGCCTCCCGTGAACGCGTGTCTGTGCGCGTGTCTTAC 46970  
Qy 301 AAGCCTGGGTGACCTCCTCAGCAGCTGCGAGCTCTCTGTCAAGCTGGGGTGGAGCAGGC 360  
Db 46971 AAGCCTGGGTGACCTCCTCAGCAGCTGCGAGCTCTCTGTCAAGCTGGGGTGGAGCAGGC 47030  
Qy 361 CCTGAGCAGCTGCGAGCTGCCCTCTTAACCCCTCTGCCCCCTCCACAGCTGTGTGAGAACGA 420  
Db 47031 CCTGAGCAGCTGCGAGCTGCCCTCTTAACCCCTCTGCCCCCTCCACAGCTGTGTGAGAACGA 47090  
Qy 421 CCGAAACAAGAAAGAGAGGAGTGCACCAAGCCGAGTGTCTGTAGAGCTACACCTGTAC 480  
Db 47091 CCGAAACAAGAAAGAGAGGAGTGCACCAAGCCGAGTGTCTGTAGAGCTACACCTGTAC 47150  
Qy 481 GCCGGAGTGGGGAGCTCATTGAGAGGTCGCAAAAGCCGACCAAGAAACCTTCCTCCCTGC 540  
Db 47151 GCCGGAGTGGGGAGCTCATTGAGAGGTCGCAAAAGCCGACCAAGAAACCTTCCTCCCTGC 47210  
Qy 541 CCTCTGCCAGCTGGGCAATACACTAGCGTATGGCTTTCCCGCCGCTCGAGGGTGGGAT 600  
Db 47211 CCTCTGCCAGCTGGGCAATACACTAGCGTATGGCTTTCCCGCCGCTCGAGGGTGGGAT 47270





Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M.,  
 Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
 Moser, M., Neal, D., Newton, J., Newton, J., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokenkwo, S., Oguh, M., Okwuonu, G.,  
 Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y.,  
 Rives, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I.,  
 Sodergren, E., Sonaika, T., Sparks, A., Stanley, H., Stone, K.,  
 Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
 Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
 Wu, C., Wu, Y., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G. and Gibbs, R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 162393)  
 Worley, K.C.  
 Direct Submission  
 Submitted (19-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 162393)  
 Worley, K.C.  
 Direct Submission  
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 12, 2002 this sequence version replaced gi:18701588.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Center project name: GOAF  
 Center clone name: CH230-15011  
 ----- Summary Statistics  
 : Sequencing vector: Plasmid;  
 : Chemistry: Dye-terminator Big Dye; 100% of reads  
 : Assembly program: Phrap; version 0.990329  
 : Consensus quality: 96609 bases at least Q40  
 : Consensus quality: 101360 bases at least Q30  
 : Consensus quality: 105362 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* : (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 72 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* :  
 \* 1242: contig of 1242 bp in length  
 \* 1243: gap of unknown length  
 \* 1343: contig of 1085 bp in length  
 \* 2428: gap of unknown length  
 \* 2528: contig of 1108 bp in length  
 \* 3635: gap of unknown length  
 \* 3735: contig of 1086 bp in length  
 \* 4821: gap of unknown length  
 \* 4822: contig of 1719 bp in length  
 \* 6641: gap of unknown length  
 \* 6741: contig of 1476 bp in length  
 \* 8217: gap of unknown length  
 \* 9317: contig of 1030 bp in length  
 \* 9347: gap of unknown length  
 \* 9447: contig of 1245 bp in length

10692: gap of unknown length  
 10792: contig of 1060 bp in length  
 11851: gap of unknown length  
 11951: contig of 1558 bp in length  
 13509: gap of unknown length  
 13609: contig of 1260 bp in length  
 14869: gap of unknown length  
 14969: contig of 1306 bp in length  
 16275: gap of unknown length  
 16375: contig of 1065 bp in length  
 17440: gap of unknown length  
 17540: contig of 1416 bp in length  
 18956: gap of unknown length  
 19056: contig of 1005 bp in length  
 20061: gap of unknown length  
 20161: contig of 1071 bp in length  
 21232: gap of unknown length  
 21332: contig of 1046 bp in length  
 22378: gap of unknown length  
 22478: contig of 1151 bp in length  
 23629: gap of unknown length  
 23729: contig of 1264 bp in length  
 24993: gap of unknown length  
 25093: contig of 1555 bp in length  
 26548: gap of unknown length  
 26748: contig of 1635 bp in length  
 28383: gap of unknown length  
 28483: contig of 1175 bp in length  
 29658: gap of unknown length  
 29758: contig of 1790 bp in length  
 31548: gap of unknown length  
 31648: contig of 1267 bp in length  
 32915: gap of unknown length  
 33015: contig of 1117 bp in length  
 34132: gap of unknown length  
 34232: contig of 1144 bp in length  
 35376: gap of unknown length  
 35476: contig of 1359 bp in length  
 36835: gap of unknown length  
 36935: contig of 1184 bp in length  
 38119: gap of unknown length  
 38219: contig of 1829 bp in length  
 40048: gap of unknown length  
 40148: contig of 1500 bp in length  
 41648: gap of unknown length  
 41748: contig of 2275 bp in length  
 44023: gap of unknown length  
 44123: contig of 1174 bp in length  
 45297: gap of unknown length  
 45397: contig of 1136 bp in length  
 46533: gap of unknown length  
 46633: contig of 1818 bp in length  
 48451: gap of unknown length  
 48551: contig of 1338 bp in length  
 49889: gap of unknown length  
 49989: contig of 2077 bp in length  
 52066: gap of unknown length  
 52166: contig of 1558 bp in length  
 53724: gap of unknown length  
 53824: contig of 1526 bp in length  
 55350: gap of unknown length  
 55450: contig of 2385 bp in length  
 57835: gap of unknown length  
 57935: contig of 1535 bp in length  
 59470: gap of unknown length  
 59570: contig of 1479 bp in length  
 61049: gap of unknown length  
 61149: contig of 1470 bp in length  
 62619: gap of unknown length  
 62719: contig of 2406 bp in length  
 65125: gap of unknown length  
 65225: contig of 1202 bp in length  
 66427: gap of unknown length  
 66527: contig of 1202 bp in length

	* : 6528	68713:	contig of 2186 bp in length	
*	: 68714	68813:	gap of unknown length	
*	: 68814	71416:	contig of 2603 bp in length	
*	: 71417	71516:	gap of unknown length	
*	: 71517	73660:	contig of 2144 bp in length	
*	: 73661	73760:	gap of unknown length	
*	: 73761	75002:	contig of 1742 bp in length	
*	: 75003	75602:	gap of unknown length	
*	: 75603	77965:	contig of 2363 bp in length	
*	: 77966	78065:	gap of unknown length	
*	: 78066	79509:	contig of 1444 bp in length	
*	: 79510	79609:	gap of unknown length	
*	: 79610	81649:	contig of 2040 bp in length	
*	: 81650	81749:	gap of unknown length	
*	: 81750	83575:	contig of 1826 bp in length	
*	: 83576	83675:	gap of unknown length	
*	: 83676	86277:	contig of 2602 bp in length	
*	: 86278	86377:	gap of unknown length	
*	: 86378	88701:	contig of 2324 bp in length	

  

Query Match :    20.4% ; Score 204.4 ; DB 2 ; Length 162393;  
Best Local Similarity     62.8%; Pred. No. 3.8e-34;  
Matches 453; Conservative    0; Mismatches 147; Indels 121; Gaps 4;

---

Qy	284	TGTTGGCGGTTCACAGCCTGGTGACCTCCCTCACAGCTGCAGCTCTCTGTCAAG	343
Db	97462	TGTGTGTGTGTGTGTGTGTGTGTATCTGTGTCCATCTGGTGTGGTAGATAG	97403
Qy	344	CTGGGGTGGAGGCCCTGAGCAGCTGAGCTGCCCTCTTAACCCC---CTCTGCC	400
Db	97402	CATAGGCTGGCGCTGACCCTGAACCGAAGCATTCACTCTAACCCACCATGSCACCC	97343
Qy	401	TCCACGCTGTGAGAAGACGCGAAACAAGAAAGAGGAGGTGCCAAGCCGAGNGC	460
Db	97342	CCCCAGCGGTGCNAAGCACCGAAACAAAAGAAAGAAAGAAACACCCACGCCGAGTGC	97283
Qy	461	TCTGAGAGCTACAGCTGACCGCGAGGTGGGGAGCTCATTTGAGAGGTGCGCAAAGCG	520
Db	97282	TCAGAGAGCTACAGCTGACACCGAGGTGGGGAACTCATCGAAGGTGCGCAAAGCA	97223
Qy	521	CACCAAGAAACCTTCCCTGCCCTGCCAGCTGGCAATAACATACGGTATGCTTCC	580
Db	97222	CACCAAGAGACCTTCCCGGCCCTGCGCAGCTGGCAAGTACATACGGTA-----	97172
Qy	581	CCCGSCTGTCAGSGGTGGGATTGTGCCAGGCGCACAGGGCCAGGATGGGCCCTCTCAGGC	640
Db	97171	-----	97172
Qy	641	ACCCCTTCTGTGCCAGCAGATCTCTGCTCCTCCCTTCCCTCTCTTCTCCCTCT	700
Db	97171	-----TGCCAGCCAGGCCCTCTCTGTCTATCTCCCTCTGTGCTCTACACTGT	97123
Qy	701	CCTGCTGCCCTTCCCAAGGAGCTCCCAAGAGTGAAGGCTGGGTAGAGGCGAGCCCTGT	760
Db	97122	C-----AGGAAGTAGGGNTGGGTGGAGG	97098
Qy	761	GGGGCTGGAGCAGGCTGAGAAGGGGTGCCATGGAGAAGGCCCTCACTCTCCCTCC	820
Db	97097	GACGAGGGGCCAGGCTGAACGGGTGTC----GTAGCTGAGGCCCTTACTCTCCCTGCC	97041
Qy	821	TCCCCAGAAACAACAGCTCAGAACAGCTGTCTCTCGACATTGACCTCTGGGACAAGT	880
Db	97040	TCCCCAGAACACAGCTCAGAACACGTGTCTCTCTGGACATTGACCTCTGGGACAAGT	96981
Qy	881	TCAGTGAACCTCCACCAAGTGCATATTAGACTGTGGAGTTGCCAAGCAGCTGCCCG	940
Db	96980	TCAGTGAACCTCCACCAAGTGTATCATTTAAGACTGTGGAGTTGCCAAGCAGCTTCCCG	96921
Qy	941	GCTTCACACCCCTCACCATCGCCAGCACAGATCACCCCTCTCAAGGCTGCCTGCCTGACA	1000
Db	96920	GCTTCACACCCCTCACCATTGGCGACAGATTACCTTCTCAAGGCTGCCTGCCTGACA	96861
Qy	1001	T	1001

Db 96860 t 96860

RESULT 10  
AR052138 704 bp DNA linear PAT 29-SEP-1999  
LOCUS Sequence 5 from patent US 5830760;  
DEFINITION AR052138  
ACCESSION AR052138  
VERSION AR052138.1 GI:5975502  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE  
1 (bases 1 to 704)  
AUTHORS Tsai,S. and Collins,S.J.  
TITLE Creating novel hematopoietic cell lines by expressing altered retinoic acid receptors  
JOURNAL Patent: US 5830760-A 5 03-NOV-1998;  
FEATURES Location/Qualifiers  
source 1..704  
/organism="unknown"  
BASE COUNT 150 a 232 c 202 g 120 t  
ORIGIN

Query Match 17.4%; Score 174; DB 6; Length 704;  
Best Local Similarity 100.0%; Pred. No. 2.7e-27;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACAAACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAGTGA 887  
Db 41 GAACAACAGCTCAGAACAAACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAGTGA 100

QY 888 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTGTGCCAAGCAGCTGCCCGGCTTCAC 947  
Db 101 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTGTGCCAAGCAGCTGCCCGGCTTCAC 160

QY 948 CACCTCACCATCGCCGACAGATCACCCCTCTCAAGGCTGCCTGGGACAT 1001  
Db 161 CACCTCACCATCGCCGACAGATCACCCCTCTCAAGGCTGCCTGGGACAT 214

RESULT 11  
S50916  
LOCUS S50916 1284 bp mRNA linear PRI 29-JUN-2000  
DEFINITION PML-RAR fusion gene [fusion transcript] (human, mRNA Partial, 1284 nt).  
ACCESSION S50916  
VERSION S50916.1 GI:234248  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.  
REFERENCE 1 (bases 1 to 1284)  
AUTHORS de The,H., Lavau,C., Marchio,A., Chomienne,C., Degos,L. and Dejean,A.  
TITLE The PML-RAR alpha fusion mRNA generated by the t(15;17) translocation in acute promyelocytic leukemia encodes a functionally altered RAR  
JOURNAL Cell 66 (4), 675-684 (1991)  
MEDLINE 91347369  
PUBMED 1652369  
REMARK GenBank staff at the National Library of Medicine created this entry [NCBI gibbsq 50916] from the original journal article.  
This sequence comes from fig2b.  
FEATURES Location/Qualifiers  
source 1..1284  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
gene <1..1284  
CDS /gene="PML-RAR fusion gene"  
<1..1251

```

/gene="PML-RAR fusion gene"
/note="This sequence comes from fig2b; conceptual
translation presented here differs from translation in
publication; RAR alpha"
/codon_start=1
/product="retinoic acid receptor alpha"
/protein_id="AABL9602.2"
/db_xref="GI:881561"
/translation="PSNSHVASGAGAAIETOSSSEIIVPSPPPPLRIYKPCFV
CQKSSGYHYGASACGCGFFRRSGIKNMVYTRCHKDKNCIINKVNRQCQYCRLOKQ
FVYMGSKSVNRDNRRKKEVPEKCESYTLTPEKVEIKTKVKAHQETFPALCOAG
KYITNNSBQVRSLDLDMDKFSSELSTKCIITKTFEAKOLPGCTTLTIDAOITLLKAA
CDLILRLCTRVTYPPQDTWTFSDGLTINRTQMHNAGFGLDLYAFANQLLPLEMD
DABTGSLASLCAGORQDEQDPQVMDLQEPLEALKVTYVRKRFRSRHMFPMMLMK
ITDLRSISAKGAERVITLKEWIPSPSPPLLIQEMLENSGLDTLSGQPGGGRDGGGLA
PPFGSCPSLSPSSNSRTHPS"
285 a 426 c 371 g 202 t

```

	Query Match	17.4%	Score 174;	DB 9;	Length 1284;	
	Best Local Similarity	100.0%;	Pred. No. 2.5e-27;			
	Matches 174;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;	
Qy	828	GAACAACAGCTCAGACAACAGTGTCTCTCTGGACATTGACCTCTGGGACAAGTTTCAGTGA	887			
Db	492	GAACAACAGCTCAGACAACAGTGTCTCTCTGGACATTGACCTCTGGGACAAGTTTCAGTGA	551			
Qy	888	ACTCTCCACCAAGTGCATCAATTAAGACTGTGGAGTTGCGCAACGACGTCGCCGGCTTTCAC	947			
Db	552	ACTCTCCACCAAGTGCATCAATTAAGACTGTGGAGTTGCGCAACGACGTCGCCGGCTTTCAC	611			
Qy	948	CACCTTCACCATCCCGACACAGATCACCTCTCTCAAGGCTGCCTGGACAT	1001			
Db	612	CACCTTCACCATCCCGACACAGATCACCTCTCTCAAGGCTGCCTGGACAT	665			

```

RESULT 12
AF242868      1572 bp      mRNA      linear      SYN 24-APR-2000
LOCUS        Synthetic construct ECFP-retinoic acid receptor ligand binding
DEFINITION   domain fusion protein mRNA, complete cds.

ACCESSION   AF242868
VERSION      AF242868.1  GI:7638252
KEYWORDS     .
SOURCE       synthetic construct.
ORGANISM     artificial sequences.
1. (bases 1 to 1572)
Llopis,J., Westin,S., Ricote,M., Wang,J., Cho,C.Y., Kurokawa,R.,
Mullen,T.M., Rose,D.W., Rosenfeld,M.G., Tsien,R.Y. and Glass,C.K.
Ligand-dependent interactions of coactivators steroid receptor
coactivator-1 and peroxisome proliferator-activated receptor
binding protein with nuclear hormone receptors can be imaged in
live cells and are required for transcription
Proc. Natl. Acad. Sci. U.S.A. 97 (8), 4363-4368 (2000)
20226109
JOURNAL      10760302
MEDLINE      2
PUBMED       2 (bases 1 to 1572)
AUTHORS      Llopis,J., Westin,S., Ricote,M., Wang,J., Cho,C.Y., Kurokawa,R.,
Mullen,T.M., Rose,D.W., Rosenfeld,M.G., Tsien,R.Y. and Glass,C.K.
TITLE        Direct Submission
SUBMITTED    (09-MAR-2000) Medicine/Cellular and Molecular Medicine,
UC San Diego, 9500 Gilman Drive, La Jolla, CA 92103, USA
FEATURES     Location/Qualifiers
1. .1572
   /organism="synthetic construct"
   /db_xref="taxon:32630"
   /focus
   685. .1572
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
1. .1572
   /codon_start=1

source
source
CDS

```

/transl_table=11	
/product="ECSP-retinoic acid receptor ligand binding domain fusion protein"	
/protein_id="AAF65453.1"	
/db_xref="GI:7638253"	
/translation="MYSKGELSTGVVPIILYELDGVNGHRSVSGEGGDATYGLKLT LKFTCTTGKVPWPVLTITLTGMGVCFSRYPDHMKQHDFFKSAMPEGGYVOERTIFFK DGNYKTRAEVKFEGDITLVNRIELKIGDFKEDNGLGHKLEYINYISHNVYITADKQKN GKNAHFTRAEKIDEGSQLADHYQONTPIGDGPVLLPDNHYLTSQSALSQDPNKRHD MWLEFVTAARMPKPCSESYTLTPVGLIEKVRKAHQETFPALCOLGKYTTNNNSE ORYSLDIDLWKTSELSKTCIIKTVEFAKOLPGFTTLTIAQDTLLKAACLDLILIRI CTRYTPEQDPTWTSDDGLTLNRTOMHNRAGFPLDLYFAFANQLLPLEMDDAEATGCLLSA ICLIGDQDLEQFDVDMQLQEPLEALKVYVRKRRFSRPHMFPMKMLKITDLKRSIA KSPSSRRITLMEIPGSMPLLIQIEMENSEGLDTLSGQPGGGRDGGGLAPPDSCGSPSS LSPSSRRITLMEIPGSMPLLIQIEMENSEGLDTLSGQPGGGRDGGGLAPPDSCGSPSS	
BASE COUNT	346 a 536 c 456 g 234 t
ORIGIN	
Query Match	17.4% ; Score 174 ; DB 12 ; Length 1572 ;
Best Local Similarity	100.0% ; Pred. No. 2.4e-37 ;
Matches 174 ; Conservative 0 ; Mismatches 0 ; Indels 0 ; Gaps 0 ;	
Qy 828	GAACAACAGCTCAGAACAACTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 887
Db 813	GAACAACAGCTCAGAACAACTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 872
Qy 888	ACTCTCCACCAAGTGCATCATTAAGACATGTGGAGTTGCGCCAAAGCAGCTGCCCGCTTCAC 947
Db 873	ACTCTCCACCAAGTGCATCATTAAGACATGTGGAGTTGCGCCAAAGCAGCTGCCCGCTTCAC 932
Qy 948	CACCTTCACCATCGCCGACAGATCACCTCTCTCAAGGTCGCTCGCTGGACAT 1001
Db 933	CACCTTCACCATCGCCGACAGATCACCTCTCTCAAGGTCGCTCGCTGGACAT 986
RESULT 13	
HSRAR	HSRAR 1920 bp mRNA linear PRI 12-SEP-1993
LOCUS	Human mRNA for retinoic acid receptor.
DEFINITION	X06538
ACCESSION	X06538.1 GI:35873
VERSION	
KEYWORDS	DNA binding protein; hormone receptor; receptor; retinoic acid receptor.
SOURCE	Homo sapiens.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 260 to 1920)
TITLE	Petrovich,M., Brand,N.J., Krust,A. and Chambon,P.
	A human retinoic acid receptor which belongs to the family of nuclear receptors
JOURNAL	Nature 330 (6147), 444-450 (1987)
MEDLINE	88065872
PUBMED	2825025
REFERENCE	2 (bases 1 to 420)
AUTHORS	Chambon,P.
TITLE	Direct Submission
JOURNAL	Submitted (22-DEC-1988)
COMMENT	cell line=MC-F7; library=lambda gt10; clone=p63.
FEATURES	Location/Qualifiers
source	1..1920
	/organism="Homo sapiens"
CDS	/db_xref="taxon:9606"
	317..1615
	/note="retinoic acid receptor (AA 1 - 432)"

```

RESULT 13
LOCUS      HSRAR
DEFINITION Human mRNA for retinoic acid receptor.
ACCESSION X06538
VERSION    1
KEYWORDS   DNA binding protein; hormone receptor; receptor; retinoic acid
           receptor.
SOURCE     Homo sapiens
ORGANISM   Homo sapiens
REFERENCE  1 (bases 260 to 1920)
AUTHORS   Petkovich, M., Brand N.J., Krust, A. and Chambon, P.
TITLE      A human retinoic acid receptor which belongs to the family of
           nuclear receptors
JOURNAL    Nature 330 (6147), 444-450 (1987)
MEDLINE    88065872
PUBMED     2825025
REFERENCE  2 (bases 1 to 420)
AUTHORS     Chambon, P.
TITLE       Direct Submission
JOURNAL     Submitted (22-DEC-1988)
COMMENT     cell line=McF-7; library=lambda gt10; clone=p63.
FEATURES   Location/Qualifiers
            1..1920
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             317..1615
              /note="retinoic acid receptor (AA 1 - 432)"
              /codon_start=1
              /protein_id="CAA29787.1"
              /db_xref="GI:35874"
              /db_xref="SWISS-PROT:P10276"
              /translation="MLGGLSPPGALTLQLHQLPVSGYSPSPATIEQTSSSSSEIVPS
PPSPPLPRILKPCFVQDKSSGHYGVSAECGKGFRRSIQKNMVTCHRDKNCLL
VKYTRNRCQYCRQLQCFEYGMSESRVNRNDKKKKEVPKPECSYTLTPEDKGLIEK
NRKHAETPAPCOLGKYTTNNSSEQRSLDLDWKFSELSTKCIKTVEFAKOLP"

```

FTTLTIADQITLLKAAACDILILRICITRYTPEQDTWTFSDGLTLNRTOMHNAGFGPLT  
DLVAFANQLLPLEMDDAETGELLCAICLCGRQDLEQDQDVRDMLQEPLEALKAYVIR  
KRPSRPHMPKMLKMTDLSRSISAKGAERVITLTKMEIPGMPPLIQEMLENSEGLDT  
LSQPGGGGRDGGGLAPPGSCPSLSPSSNRSSPATHSP"

misc\_feature

/note="put. DNA-binding domain"

old\_sequence

/note="gc was cg in [1]"

misc\_feature

/citation=[1]

old\_sequence

/note="put. ligand-binding domain"

old\_sequence

/note="g was c in [1]"

old\_sequence

/citation=[1]

old\_sequence

/note="g was c in [1]"

old\_sequence

/citation=[1]

old\_sequence

/note="g was c in [1]"

Query Match 17.4%; Score 174; DB 9; Length 1920;  
Best Local Similarity 100.0%; Pred. No. 2.4e-27;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTGCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 887

Db 856 GAACAACAGCTCAGAACACGTGCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 915

QY 888 ACTCTCACCAGTGCATTAAGACTGTGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 947

Db 916 ACTCTCACCAGTGCATTAAGACTGTGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 975

QY 948 CACCTTCACCAATCGCCGACAGATCACCTCTCAAGGCTGCCTCGCTGGACAT 1001

Db 976 CACCTTCACCAATCGCCGACAGATCACCTCTCAAGGCTGCCTCGCTGGACAT 1029

RESULT 14

HSU41743

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

Source

CDS

/db\_xref="GI:1314309"  
/translation="MEDSDMDMSPRLPQNYLFGCELKADKDYHFVDMDENERQSLSL  
KRVSLGAGAKDELHIVAEAMNVEGSPKIVTTLATLKMVSQPTVSLGFEITPPVVLRL  
KCGSGPVHISGHOHLVALETQSSSEIEIVSPSPPLPRIYKPCFVQCDKSSGSHYGV  
SACECKGFFRRSIOKNMVTCHRDKNCLINKVRNRCQICRLQKCFEVMGSKESVRN  
DRNKKKEVPKPESESYYLTTPVEGELIEKVRKHOETFPALCOLKQYKTYNNSEQVR  
SLIDLDWKFSELSKCIITVFQKLPFTLTDVAFANQLLPLEMDDAETGLLSAICL  
YTPEDQTMFTSDGLTLNRTOMHNAGFGPLTDLVAFANQLLPLEMDDAETGLLSAICL  
ICGDRDLEQDQDVRDMLQEPLEALKAYVIRKRPSRPHMPKMLKMTDLSRSISAKGA  
ERVITLTKMEIPGMPPLIQEMLENSEGLDTLSQPGGGGRDGGGLAPPGSCPSLSP  
SSNRSSPATHSP"

/protein\_id="AAB00113.1"

/db\_xref="GI:1314310"

BASE COUNT 427 a 598 c 557 g 362 t  
ORIGIN

Query Match 17.4%; Score 174; DB 9; Length 1944;  
Best Local Similarity 100.0%; Pred. No. 2.4e-27;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTGCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 887

Db 887 GAACAACAGCTCAGAACACGTGCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 946

QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 947

Db 947 ACTCTCCACCAAGTGCATTAAGACTGTGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 1006

QY 948 CACCTTCACCAATCGCCGACAGATCACCTCTCAAGGCTGCCTCGCTGGACAT 1001

Db 1007 CACCTTCACCAATCGCCGACAGATCACCTCTCAAGGCTGCCTCGCTGGACAT 1060

RESULT 15

AK098172

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

AUTHORS

TITLE

JOURNAL

COMMENT

AK098172

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

Homo sapiens

```

FEATURES      RAB; annotation: HRI and RAB.
SOURCE        Location/Qualifiers
              1..2008
              /organism="Homo sapiens"
              /db_xref="taxon:9606"
              /clone="TRACH2015486"
              /tissue_type="trachea"
              /clone_lib="TRACH2"
              /note="cloning vector: pME18SFL3"

BASE COUNT   405 a 669 c 576 g 358 t
ORIGIN

Query Match      17.4%; Score 174; DB 9; Length 2008;
Best Local Similarity 100.0%; Pred. No. 2.4e-27;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 828 GAACAACAGCTCAGAACACGCTCTCTCTGGACATTTGACCTCTGGGACAAAGTTTCAGTGA 887
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 680 GAACAACAGCTCAGAACACGCTCTCTCTGGACATTTGACCTCTGGGACAAAGTTTCAGTGA 739

Oy 888 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCCGCCAAGCAGCTGCCCGGCTTCAC 947
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 740 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCCGCCAAGCAGCTGCCCGGCTTCAC 799

Oy 948 CACCCCTCACCATCGCCGACCAAGATCACCTCTCAAGGCTGCCTGGACAT 1001
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 800 CACCCCTCACCATCGCCGACCAAGATCACCTCTCAAGGCTGCCTGGACAT 853

```

Search completed: March 30, 2003, 13:00:12  
Job time : 2850.53 secs



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 07:57:38 ; Search time 196.299 Seconds  
(without alignments)  
11483.757 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_10000\_11000  
Perfect score: 1001  
Sequence: 1 cccagggagactgcagctgg.....aaggctgcctgcctggacat 1001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

1:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20:	/SID22/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24:	/SID22/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1001	100.0	20512	24	Genomic DNA encodi
2	174	17.4	1481	17	RAR-alpha RNA sequ
3	174	17.4	1481	17	RAR-alpha DNA sequ
4	174	17.4	1692	23	DNA encoding novel
5	174	17.4	1944	23	DNA encoding novel
6	174	17.4	1944	23	DNA encoding novel
7	174	17.4	2073	23	DNA encoding novel
8	174	17.4	2086	24	CDNA encoding huma
9	174	17.4	2457	23	DNA encoding novel

10	174	17.4	2658	16	AAO81477
11	174	17.4	2907	24	ABK84517
12	174	17.4	2907	24	ABL65214
13	174	17.4	2928	13	AAO29338
14	174	17.4	2928	20	AAV64991
15	174	17.4	2940	10	AA90124
16	174	17.4	2940	16	AAO81476
17	174	17.4	3036	13	AAQ29334
18	174	17.4	3036	19	AAV20474
19	174	17.4	3036	20	AAV64990
20	174	17.4	3052	23	AA83047
21	174	17.4	3511	17	AA833246
22	174	17.4	3511	17	AA833259
23	174	17.4	3511	21	AA83656
24	156.6	15.6	2930	23	AA83048
25	156	15.6	301	22	AA879992
26	140.8	14.1	456	23	AA83044
27	117.6	11.7	1375	24	ABK37492
28	117.6	11.7	1399	24	ABK37493
29	117.6	11.7	8887	24	ABK37494
30	117.6	11.7	9151	24	ABK37495
31	112.8	11.3	1351	16	AAO79935
32	112.8	11.3	1912	12	AAO10388
33	112.8	11.3	1913	16	AAO5202
34	112.8	11.3	2521	13	AAQ20542
35	112.8	11.3	2740	13	AAQ20541
36	112.8	11.3	2989	10	AA900093
37	112.8	11.3	2992	16	AAO79934
38	112.8	11.3	2992	17	AAO6491
39	111.2	11.1	2992	15	AAQ55572
40	109.2	10.9	1576	12	AAQ10226
41	109.2	10.9	1903	12	AAQ10389
42	109.2	10.9	1903	16	AAO5196
43	107.2	10.7	1416	23	AA86580
44	102.4	10.2	482	23	AA81232
45	102.4	10.2	5838	22	AAH98429

## ALIGNMENTS

RESULT 1  
AAL38339  
ID AAL38339 standard; DNA; 20512 BP.  
XX AC AAL38339;  
XX 15-AUG-2002 (first entry)  
XX Genomic DNA encoding the human nuclear hormone receptor protein.  
XX Human nuclear hormone receptor; therapeutic agent; gene therapy;  
XX Immune response; chromosome 17; single nucleotide polymorphism;  
XX gene; ds.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX CDS 1684..15208  
FT FT /\*tag= a  
FT FT /product= "Human nuclear hormone receptor protein"  
FT FT 1684..1846  
FT FT /\*tag= b  
FT FT /number= 1  
FT FT 1847..6808  
FT FT /\*tag= c  
FT FT /number= 1  
FT FT 6809..6957  
FT FT /\*tag= d  
FT FT /number= 2  
FT FT 6958..8277  
FT FT /\*tag= e  
FT FT /number= 2

FT exon 8278..8419  
 FT /\*tag= f  
 FT /number= 3  
 FT intron 8420..10406  
 FT /\*tag= g  
 FT /number= 3  
 FT exon 10407..10567  
 FT /\*tag= h  
 FT /number= 4  
 FT intron 10568..10827  
 FT /\*tag= i  
 FT /number= 4  
 FT exon 10828..11004  
 FT /\*tag= j  
 FT /number= 5  
 FT intron 11005..12798  
 FT /\*tag= k  
 FT /number= 5  
 FT exon 12799..13003  
 FT /\*tag= l  
 FT /number= 6  
 FT intron 13004..13759  
 FT /\*tag= m  
 FT /number= 6  
 FT exon 13760..13918  
 FT /\*tag= n  
 FT /number= 7  
 FT intron 13919..14505  
 FT /\*tag= o  
 FT /number= 7  
 FT exon 14506..14658  
 FT /\*tag= p  
 FT /number= 8  
 FT intron 14659..15143  
 FT /\*tag= q  
 FT /number= 8  
 FT exon 15144..15208  
 FT /\*tag= r  
 FT /number= 9  
 FT allele replace(4084,C)  
 FT allele replace(6482,A)  
 FT allele replace(8066,G)  
 FT allele replace(8699,C)  
 FT allele replace(12897,T)  
 FT allele replace(14442,C)  
 FT /\*tag= x  
 XX W0200231146-A2.  
 PN 18-APR-2002.  
 XX 05-OCT-2001; 2001WO-US31095.  
 XX 11-OCT-2000; 2000US-239117P.  
 XX 19-OCT-2000; 2000US-0691220.  
 XX (PEKE ) PE CORP NY.  
 PA  
 PI Wei M, Ye J, Yan C, Ketchum KA, Di Francesco V, Beasley EM;  
 XX WPI: 2002-426282/45.  
 DR P-PSDB; AAO21489.  
 DR  
 XX New human nuclear hormone receptor proteins and nucleic acids, useful  
 PT as models or targets for developing human therapeutic targets, and in  
 PT identifying therapeutic proteins and modulators of nuclear hormone  
 PT receptor expression  
 PT

PS Claim 1; Fig 3; 73pp; English.

XX The invention relates to an isolated peptide of a novel human nuclear  
 CC hormone receptor with a fully defined sequence of 457 amino acids given  
 CC in the specification. The novel human nuclear hormone receptor peptides  
 CC and nucleic acids encoding them can be used as models for the development  
 CC of human therapeutic targets, aid in the identification of therapeutic  
 CC proteins, and serve as targets for the development of human therapeutic  
 CC agents that modulate nuclear hormone receptor activity in cells and  
 CC tissues that express the nuclear hormone receptor. The nucleic acids may  
 CC be used as a query sequence to perform searches against sequence  
 CC databases to identify family members or related sequences, as probes or  
 CC primers, to construct recombinant vectors, to identify compounds that  
 CC modulate nuclear hormone receptor nucleic acid expression, in gene  
 CC therapy, and as antisense constructs to control nuclear hormone receptor  
 CC gene expression in cells, tissues or organisms. The polypeptides can be  
 CC used to raise antibodies or to elicit an immune response, as a reagent in  
 CC assays designed to determine protein levels in biological fluids, as  
 CC markers for tissues in which a corresponding protein is expressed, to  
 CC identify a binding partner/ligand to develop a system for the  
 CC identification of inhibitors of the binding reaction, in drug screening  
 CC assays, and to identify compounds that modulate protein activity. This  
 CC polynucleotide sequence represents the genomic DNA encoding the human  
 CC nuclear hormone receptor protein of the invention.

XX SQ Sequence 20512 BP; 3591 A; 5987 C; 6281 G; 4541 T; 112 other;

Query Match 100.0%; Score 1001; DB 24; Length 20512;  
 Best Local Similarity 100.0%; Pred. No. 5.3e-223;  
 Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 CCCAGGGAGCTGCAGCTGGGAGGCGTGGTGAGTGGAGGCGGGAGAGGACTTCCTGG 60  
 Db 10000 CCCAGGGAGCTGCAGCTGGGAGGCGTGGTGAGTGGAGGCGGGAGAGGACTTCCTGG 10059  
 QY 61 GGAAGAGGAGGAGCAGACCTAGGAGGSCACCGTCGCTGAGTGGAGTGGAGTGA 120  
 Db 10060 GGAAGAGGAGGAGCAGACCTAGGAGGSCACCGTCGCTGAGTGGAGTGGAGTGA 10119  
 QY 121 CGGTGGGGATAGCATCGGCTGGCTATGGGGTGGGTGGGGGGTGTGTGCAGGGCCAC 180  
 Db 10120 CGGTGGGGATAGCATCGGCTGGCTATGGGGTGGGTGGGGGGTGTGTGCAGGGCCAC 10179  
 QY 181 AGCTGTGCTCATGGGCTTCCTGGGCAGAACTTGATGTGGGTGGGTGGGCATGGAGG 240  
 Db 10180 AGCTGTGCTCATGGGCTTCCTGGGCAGAACTTGATGTGGGTGGGTGGGCATGGAGG 10239  
 QY 241 GCTGGAGTGGTGGCAATGCCTTGCTGCCGCTGAACGCGTCTGTGTGCGGTGCTTAC 300  
 Db 10240 GCTGGAGTGGTGGCAATGCCTTGCTGCCGCTGAACGCGTCTGTGTGCGGTGCTTAC 10299  
 QY 301 AAGCCTGGGTGACCTCTCAGCAGCTGGCAGCTCTCTCTCAGGCTGGGGGTGGAGGAGC 360  
 Db 10300 AAGCCTGGGTGACCTCTCAGCAGCTGGCAGCTCTCTCTCAGGCTGGGGGTGGAGGAGC 10359  
 QY 361 CCTGAGCAGCTGCAGCTGCCCTCTTAAACCCCTCTGCCCTCCACAGCTGTGAGAAACA 420  
 Db 10360 CCTGAGCAGCTGCAGCTGCCCTCTTAAACCCCTCTGCCCTCCACAGCTGTGAGAAACA 10419  
 QY 421 CCGAAACAAGAAGAAGAGGAGTGGCCCAAGCCGAGTGCCTCTGAGAGCTACACGCTGAC 480  
 Db 10420 CCGAAACAAGAAGAAGAGGAGTGGCCCAAGCCGAGTGCCTCTGAGAGCTACACGCTGAC 10479  
 QY 481 GCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGCAAGCGCACCAGGAAACCTTCCTCTGC 540  
 Db 10480 GCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGCAAGCGCACCAGGAAACCTTCCTCTGC 10539  
 QY 541 CCTCTGCCAGCTGGGCAAAATACACTACGATATGGCTTTCCCGGCTGCAAGGTGGGAT 600  
 Db 10540 CCTCTGCCAGCTGGGCAAAATACACTACGATATGGCTTTCCCGGCTGCAAGGTGGGAT 10599  
 QY 601 TTCCCGAGGCGCACAGGCGCCAGGATGGCCCTCTCAGGACACCCCTCTCTTGTGCGCAGCA 660  
 Db 10601 TTCCCGAGGCGCACAGGCGCCAGGATGGCCCTCTCAGGACACCCCTCTCTTGTGCGCAGCA 10659

[illegible]

SQ Sequence 1481 BP; 308 A; 488 C; 425 G; 260 U; 0 other;  
Query Match . 17.4%; Score 174; DB 17; Length 1481;  
Best Local Similarity 79.9%; Pred. No. 6.1e-31;  
Matches 139; Conservative 35; Mismatches 0; Indels 0; Gaps 0

---

QY	828	GAAACAACAGCTCAGAACAACGTGTCCTCTCTGGACATTGACCTCTCTGGGACAAGTTCAGTGCA	887
Dd	597	GAAACAACAGCUCAGAACAAACGUGUCUCUCUGGACAUGACCUCUGGACAAGUCCAGUCA	656
QY	888	ACTCTCCACCAAGTGCATCATTAAGACGTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC	947
Dd	657	ACUCUCCACCAAGUGCAUCAUAAGACUGUGGAGUUCGCCAAGCAGCUGC CGCGCUFCAC	716
QY	948	CACCTCACCATCGCCGACCATCACCTCTCTCAAGCTGCCTGCCTGGACAT	1001
Dd	717	CACCUCCACCAUCGCCGACCAAGACACCUCCUCAAGGUGCCUGCCUGGACAU	770

  

RESULT 3  
AAT33260  
ID AAT33260 standard; cDNA; 1481 BP.

XX	AC	AAT33260;
XX	AC	
DT	09-OCT-1996	(first entry)
XX	XX	
DE	RAR-alpha DNA sequence.	
KW	Hammerhead ribozyme; acute promyelocytic leukaemia; APL;	
KW	Lymphoma; therapy; PML-RAR-alpha; retinoic acid receptor;	
KW	external guide sequence; EGS; antisense; ss.	
OS	Homo sapiens.	
XX	XX	
PN	WO9618733-A2.	
PD	20-JUN-1996.	
XX	XX	
PF	14-DEC-1995;	95WO-US16451.
XX	XX	
PR	14-DEC-1994;	94US-0354956.
XX	XX	
PA	(INNO-) INNOVIR LAB INC.	
XX	XX	
PI	George ST, Goldberg AR, Pace U;	
XX	XX	
DR	WPI; 1996-300650/30.	
XX	XX	
PT	RNA construct(s) including ribozyme(s) and antisense	
PT	oligo:nucleotide(s) - for the inactivation of RNA associated with,	
PT	e.g. promyelocytic leukaemia or follicular lymphoma	
XX	XX	
PS	Example 1; Page 52-53; 81pp; English.	
XX	XX	
CC	Acute promyelocytic leukaemia (APL) is associated with a	
CC	translocation between the long arms of chromosomes 15 and 17,	
CC	resulting in the fusion between the retinoic acid receptor gene	
CC	(RAR-alpha, AAT33260) and PML, a putative transcription factor gene.	
CC	The fusion product RNA (see also AAT33246) junction region (AAT33251)	
CC	is the target for ribozymes (AAT33245, AAT33248, AAT33250 and AAT33252),	
CC	external guide sequences (AAT33253-54) and antisense constructs	
CC	(AAT33255-57) that specifically cleave the PML-RAR-alpha fusion	
CC	mRNA but not wild-type RAR-alpha mRNA (see also AAT33247).	
XX	XX	

  

SQ Sequence 1481 BP; 309 A; 487 C; 425 G; 257 T; 3 U; 0 other;  
Query Match 17.4%; Score 174; DB 17; Length 1481;  
Best Local Similarity 100.0%; Pred. No. 6.1e-31;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0

---

QY	828	GAAACAACAGCTCAGAACAACGTGTCCTCTGGACATTGACCTCTCTGGGACAAGTTCAGTGCA	887

Db 597 GAACAACAGCTCAGAACACGTGCTCTCTGGACATTGACCTCTGGGACAAGTTTCAGTGA 656  
QY 888 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGACAGCTGCCCGGCTTCAC 947  
Db 657 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGACAGCTGCCCGGCTTCAC 716  
QY 948 CACCTCACCATCCCGCAGCAGATCACCTCTCAAGCTGCCCTGGGACAT 1001  
Db 717 CACCTCACCATCCCGCAGCAGATCACCTCTCAAGCTGCCCTGGGACAT 770

RESULT 4  
AAS83808  
ID AAS83808 standard; cDNA; 1692 BP.  
XX AAS83808;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #19612.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG19621.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID No 19612; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1692 BP; 414 A; 470 C; 492 G; 316 T; 0 other;

Query Match 17.4%; Score 174; DB 23; Length 1692;  
Best Local Similarity 100.0%; Pred. No. 6.3e-31;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 828 GAACAACAGCTCAGAACACGTGCTCTCTGGACATTGACCTCTGGGACAAGTTTCAGTGA 887  
Db 933 GAACAACAGCTCAGAACACGTGCTCTCTGGACATTGACCTCTGGGACAAGTTTCAGTGA 992  
QY 888 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGACAGCTGCCCGGCTTCAC 947  
Db 993 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGACAGCTGCCCGGCTTCAC 1052  
QY 948 CACCTCACCATCCCGCAGCAGATCACCTCTCAAGCTGCCCTGGGACAT 1001  
Db 1053 CACCTCACCATCCCGCAGCAGATCACCTCTCAAGCTGCCCTGGGACAT 1106  
RESULT 5  
AAS83045  
ID AAS83045 standard; cDNA; 1944 BP.  
XX  
AC AAS83045;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #18849.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
OS Homo sapiens.  
XX  
PN WO200175067-A2.  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
XX WPI; 2001-639362/73.  
DR P-PSDB; ABG18858.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1; SEQ ID No 18849; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 1692 BP; 414 A; 470 C; 492 G; 316 T; 0 other;

CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 1944 BP; 427 A; 598 C; 557 G; 362 T; 0 other;  
Query Match 17.4%; Score 174; DB 23; Length 1944;  
Best Local Similarity 100.0%; Pred. No. 6.5e-31;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTCTCTCTGGACATTCACCTCGGGACAAAGTTTCAGTGA 887  
Db 887 GAACAACAGCTCAGAACACGTCTCTCTGGACATTCACCTCGGGACAAAGTTTCAGTGA 946  
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 947  
Db 947 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 1006  
QY 948 CACCCCTCACCATCGCCGACAGATCACCTCTCAAGGCTGCCCTGGACAT 1001  
Db 1007 CACCCCTCACCATCGCCGACAGATCACCTCTCAAGGCTGCCCTGGACAT 1060

RESULT 6  
AAS85136  
ID AAS85136 standard; cDNA; 1944 BP.  
XX AC AAS85136;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #20940.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG20949.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1; SEQ ID NO 20940; 103pp; English.

XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.

CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX  
SQ Sequence 1944 BP; 427 A; 598 C; 557 G; 362 T; 0 other;  
Query Match 17.4%; Score 174; DB 23; Length 1944;  
Best Local Similarity 100.0%; Pred. No. 6.5e-31;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTCTCTCTGGACATTCACCTCTGGGACAAAGTTTCAGTGA 887  
Db 887 GAACAACAGCTCAGAACACGTCTCTCTGGACATTCACCTCTGGGACAAAGTTTCAGTGA 946  
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 947  
Db 947 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 1006  
QY 948 CACCCCTCACCATCGCCGACAGATCACCTCTCAAGGCTGCCCTGGACAT 1001  
Db 1007 CACCCCTCACCATCGCCGACAGATCACCTCTCAAGGCTGCCCTGGACAT 1060

RESULT 7  
AAS85135  
ID AAS85135 standard; cDNA; 2073 BP.  
XX AC AAS85135;  
XX  
DT 13-FEB-2002 (first entry)  
XX  
DE DNA encoding novel human diagnostic protein #20939.  
XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX  
XX Homo sapiens.  
XX  
PN WO200175067-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
DR WPI; 2001-639362/73.  
DR P-PSDB; ABG20948.  
XX  
XX New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity -  
XX  
PS Claim 1; SEQ ID NO 20939; 103pp; English.

XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques

CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS94197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 2073 BP; 465 A; 612 C; 605 G; 391 T; 0 other;

Query Match 17.4%; Score 174; DB 23; Length 2073;  
Best Local Similarity 100.0%; Pred. No. 6.6e-31;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 828 GAACAACAGCTCAGAACAGCTGCTCTCTGGACATTGACCTCTGGGACAAGTTCAGTGA 887  
Db 1016 GAACAACAGCTCAGAACAGCTGCTCTCTGGACATTGACCTCTGGGACAAGTTCAGTGA 1075  
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 947  
Db 1076 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 1135  
QY 948 CACCTCACCATCGCCGACAGATCACCCTCTCAAGCTGCTGCTGGGACAT 1001  
Db 1136 CACCTCACCATCGCCGACAGATCACCCTCTCAAGCTGCTGCTGGGACAT 1189

RESULT 8  
AAL38338  
ID AAL38338 standard; cDNA; 2086 BP.  
XX AAL38338;  
XX 15-AUG-2002 (first entry)  
XX cDNA encoding human nuclear hormone receptor protein.  
XX Human nuclear hormone receptor; therapeutic agent; gene therapy;  
XX Immune response; gene; ss.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
XX CDS 127..1500  
XX /\*tag= a  
XX /product= "Human nuclear hormone receptor protein"  
XX WO200231146-A2.

XX 18-APR-2002.  
XX 05-OCT-2001; 2001WO-US31095.  
XX 11-OCT-2000; 2000US-239117P.  
XX 19-OCT-2000; 2000US-0691220.  
XX (PEKE ) PE CORP NY.  
XX Wei M, Ye J, Yan C, Ketchum KA, Di Francesco V, Beasley EM;  
XX WPI; 2002-426282/45.  
XX P-PSDB; AAO21489.  
XX New human nuclear hormone receptor proteins and nucleic acids, useful  
XX as models or targets for developing human therapeutic targets, and in  
PT

PT identifying therapeutic proteins and modulators of nuclear hormone  
PT receptor expression  
XX  
PS Claim 1; Fig 1; 73pp; English.

XX The invention relates to an isolated peptide of a novel human nuclear  
CC hormone receptor with a fully defined sequence of 457 amino acids given  
CC in the specification. The novel human nuclear hormone receptor peptides  
CC and nucleic acids encoding them can be used as models for the development  
CC of human therapeutic targets, aid in the identification of therapeutic  
CC proteins, and serve as targets for the development of human therapeutic  
CC agents that modulate nuclear hormone receptor activity in cells and  
CC tissues that express the nuclear hormone receptor. The nucleic acids may  
CC be used as a query sequence to perform searches against sequence  
CC databases to identify family members or related sequences, as probes or  
CC primers, to construct recombinant vectors, to identify compounds that  
CC modulate nuclear hormone receptor nucleic acid expression, in gene  
CC therapy, and as antisense constructs to control nuclear hormone receptor  
CC gene expression in cells, tissues or organisms. The polypeptides can be  
CC used to raise antibodies or to elicit an immune response, as a reagent in  
CC assays designed to determine protein levels in biological fluids, as  
CC markers for tissues in which a corresponding protein is expressed, to  
CC identify a binding partner/ligand to develop a system for the  
CC identification of inhibitors of the binding reaction, in drug screening  
CC assays, and to identify compounds that modulate protein activity. This  
CC polynucleotide sequence represents the cDNA encoding the human nuclear  
CC hormone receptor protein of the invention.

XX SQ Sequence 2086 BP; 437 A; 685 C; 583 G; 381 T; 0 other;

Query Match 17.4%; Score 174; DB 24; Length 2086;  
Best Local Similarity 100.0%; Pred. No. 6.6e-31;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 828 GAACAACAGCTCAGAACAGTGTCTCTCTGGACATTGACCTCTGGGACAAGTTCAGTGA 887  
Db 741 GAACAACAGCTCAGAACAGTGTCTCTCTGGACATTGACCTCTGGGACAAGTTCAGTGA 800  
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 947  
Db 801 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 860  
QY 948 CACCTCACCATCGCCGACAGATCACCCTCTCAAGCTGCTGCTGGGACAT 1001  
Db 861 CACCTCACCATCGCCGACAGATCACCCTCTCAAGCTGCTGCTGGGACAT 914

RESULT 9  
AAS83049  
ID AAS83049 standard; cDNA; 2457 BP.  
XX AAS83049;  
XX 13-FEB-2002 (first entry)  
XX DNA encoding novel human diagnostic protein #18853.

XX Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder; ss.  
XX Homo sapiens.  
XX WO200175067-A2.  
XX 11-OCT-2001.  
XX 30-MAR-2001; 2001WO-US08631.  
XX 31-MAR-2000; 2000US-0540217.  
XX 23-AUG-2000; 2000US-0649167.  
XX (HYSE-) HYSEQ INC.

PI Drmanac RT, Liu C, Tang YT;  
XX WPI: 2001-639362/73.  
DR P-PSDB; ABG18862.  
XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity  
XX  
PS Claim 1: SEQ ID No 18853; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC responsible for genetic disorders or other traits to assess biodiversity  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. AAS64197-AAS94564 represent novel human  
CC diagnostic coding sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 2457 BP; 577 A; 687 C; 700 G; 493 T; 0 other;  
Query Match 17.4%; Score 174; DB 23; Length 2457;  
Best Local Similarity 100.0%; Pred. No. 6.8e-31;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 828 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCACTGA 887  
DB 1400 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCACTGA 1459  
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAACAGCTGCCCGGCTTCAC 947  
DB 1460 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAACAGCTGCCCGGCTTCAC 1519  
QY 948 CACCTCTCACCATCGCGGACCAAGTGCATTAAGACTGTGGAGTTGCGCAACAGCTGCCCGGCTTCAC 1001  
DB 1520 CACCTCTCACCATCGCGGACCAAGTGCATTAAGACTGTGGAGTTGCGCAACAGCTGCCCGGCTTCAC 1573  
RESULT 10  
AAQ81477  
ID AAQ81477 standard; cDNA to mRNA; 2658 BP.  
XX  
AC AAQ81477;  
XX  
DT 01-SEP-1995 (first entry)  
XX  
DE RAR-alpha-403 dominant negative.  
XX  
KW RAR-alpha; retinoic acid receptor alpha; hematopoietic; stem cell;  
KW differentiation; dominant negative; retrovirus; vector; neutrophil;  
KW monocyte; mast cell; basophil; ds.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 103..1312  
FT /\*tag= a  
FT /note= "deletion of bases 1311-1596 of RAR-alpha

FT CDNA (AAQ81476) and insertion of stop codon  
XX TAG at deletion site (1312-1314)."  
PN WO9504143-A.  
XX  
XX 09-FEB-1995.  
XX  
XX 28-JUL-1994; 94WO-US08450.  
XX  
XX 28-JUL-1993; 93US-0099242.  
XX  
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
XX  
XX Collins SJ, Tsai S;  
XX  
XX WPI: 1995-082227/11.  
XX P-PSDB; AAR68024.  
XX  
XX New hematopoietic stem cell lines with specific differentiation  
XX properties - made by transfecting stem cells with nucleic acid  
XX encoding dominant negative suppressor of the retinoic acid  
XX receptor alpha, useful e.g. for hematopoietic reconstitution  
XX  
XX Disclosure; Page 58-59; 100pp; English.  
XX  
XX An RAR-alpha cDNA (given in AAQ81477) contained a truncation of  
XX sequences coding for the C-terminal 59 amino acids and part the  
XX 3' UTR of the wild-type human sequence (AAQ81476). This truncated  
XX cDNA, designated RAR-alpha-403, encoded a protein (AAR68024)  
XX containing the N-terminus, DNA-binding domain and part of the  
XX hormone-binding domain of RAR-alpha (AAR68023). RAR-alpha-403 has  
XX negative suppressor activity. Introduction into multipotent IL-3  
XX dependent FDCP mix A4 cells (murine hematopoietic) caused a  
XX switch from neutrophil/monocyte differentiation to basophil/mast  
XX cells (30-50%), with only 2% neutrophils.  
XX  
SQ Sequence 2658 BP; 550 A; 889 C; 711 G; 508 T; 0 other;  
Query Match 17.4%; Score 174; DB 16; Length 2658;  
Best Local Similarity 100.0%; Pred. No. 7e-31;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 828 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCACTGA 887  
DB 732 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCACTGA 791  
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAACAGCTGCCCGGCTTCAC 947  
DB 792 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAACAGCTGCCCGGCTTCAC 851  
QY 948 CACCTCTCACCATCGCGGACCAAGTGCATTAAGACTGTGGAGTTGCGCAACAGCTGCCCGGCTTCAC 1001  
DB 852 CACCTCTCACCATCGCGGACCAAGTGCATTAAGACTGTGGAGTTGCGCAACAGCTGCCCGGCTTCAC 905  
RESULT 11  
ABK84517  
ID ABK84517 standard; cDNA; 2907 BP.  
XX  
AC ABK84517;  
XX  
DT 14-AUG-2002 (first entry)  
XX  
XX Human cDNA differentially expressed in granulocytic cells #1088.  
XX  
KW Human; ss; granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.







PA (AVAL-) AVALON PHARM.  
XX Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX WPI: 2002-188264/24.  
XX Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set -  
XX  
PS Claim 1: SEQ ID 3551; 44pp; English.  
XX  
CC The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC oesophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilm's tumour.  
XX  
SQ Sequence 2907 BP; 561 A; 1011 C; 789 G; 546 T; 0 other;  
  
Query Match 17.4%; Score 174; DB 24; Length 2907;  
Best Local Similarity 100.0%; Pred. No. 7.1e-31;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTCAAGTGA 887  
DB 732 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTCAAGTGA 791  
  
OY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAACGACGTGCCCGGCTTCAC 947  
DB 792 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAACGACGTGCCCGGCTTCAC 851  
  
OY 948 CACCCTCACCATTGCGCGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 1001  
DB 852 CACCCTCACCATTGCGCGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 905  
  
RESULT 13  
AAQ29338  
ID AAQ29338 standard; cDNA; 2928 BP.  
XX  
AC AAQ29338;  
XX  
DT 09-MAR-1993 (first entry)  
XX  
DE RAR-alpha gene.  
XX  
KW Retinolic acid receptor; RAR-alpha; myl; acute promyelocytic leukemia;  
KW APL; translocation; chromosome 17; chromosome 15; PCR; primer;  
KW [t(15;17)(q21;q11-22)]; breakpoint; polymerase chain reaction; ss.  
XX  
OS Synthetic.  
XX  
FH Key Location/Qualifiers  
FT CDS 103..1488  
FT /\*tag= a  
XX  
PN W09216660-A.  
XX  
PD 01-OCT-1992.

XX 23-MAR-1992; 92WO-US02320.  
XX  
XX 22-MAR-1991; 91US-0673838.  
PR 22-MAR-1991; 91US-0675084.  
XX  
XX (SLOK ) SLOAN KETTERING INST CANCER.  
XX  
XX Dmitrovsky E, Evans RM, Frankel S, Kazizuka A, Miller WH;  
PI Warrell RP;  
XX  
XX WPI: 1992-349240/42.  
DR P-PSDB; AAR27534.  
XX  
XX Marker for acute promyelocytic leukaemia and other neoplasias -  
PT comprising nucleic acid and encoded abnormal retinoic acid  
PT receptor-alpha receptor  
XX  
XX Disclosure; Page 43-46; 84pp; English.  
XX  
CC The sequence given represents the nucleic acid sequence of the  
CC retinoic acid receptor (RAR)-alpha gene. This gene is disrupted in  
CC a translocation of a portion of the long arm of chromosome 17 onto  
CC the long arm of chromosome 15 [t(15;17)(q21;q11-22)]. This causes a  
CC fusion between RAR-alpha and myl which is characteristic of acute  
CC promyelocytic leukemia (APL). The breakpoint region has been cloned  
CC and it has been shown that DNA rearrangements are clustered in the  
CC region of the first intron of RAR-alpha. This sequence was isolated  
CC by polymerase chain reaction (PCR). The primers used for amplification  
CC of this sequence can also be used to amplify the translocated region.  
XX  
SQ Sequence 2928 BP; 582 A; 1012 C; 789 G; 545 T; 0 other;  
  
Query Match 17.4%; Score 174; DB 13; Length 2928;  
Best Local Similarity 100.0%; Pred. No. 7.1e-31;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
OY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTCAAGTGA 887  
DB 732 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTCAAGTGA 791  
  
OY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAACGACGTGCCCGGCTTCAC 947  
DB 792 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAACGACGTGCCCGGCTTCAC 851  
  
OY 948 CACCCTCACCATTGCGCGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 1001  
DB 852 CACCCTCACCATTGCGCGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 905  
  
RESULT 14  
AAV64991  
ID AAV64991 standard; cDNA; 2928 BP.  
XX  
AC AAV64991;  
XX  
DT 05-FEB-1999 (first entry)  
XX  
DE Human RAR-alpha cDNA.  
XX  
KW Fusion protein; myl; retinoic acid receptor-alpha; RAR; human;  
KW acute promyelocytic leukaemia; APL; t(15;17); translocation;  
KW treatment; all-trans retinoic acid; ss.  
XX  
OS Homo sapiens.  
XX  
FH Key Location/Qualifiers  
FT CDS 103..1491  
FT /\*tag= a  
FT /product= "RAR-alpha"  
FT /note= "retinoic acid receptor-alpha"  
XX  
PN US5843642-A.

```
XX 01-DEC-1998.
XX
XX 21-JUL-1993; 93US-0095728.
XX
XX 21-JUL-1993; 93US-0095728.
PR 22-MAR-1991; 91US-0673838.
PR 22-MAR-1991; 91US-0675084.
XX
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
XX Dmitrovsky E, Frankel S, Miller WH, Warrell RP;
XX WPI: 1999-044563/04.
XX P-PSDB; AAW81964.
XX
XX Diagnosis of acute promyelocytic leukaemia - by detecting nucleic
PT acid encoding abnormal retinoic acid receptor-alpha
XX
XX Disclosure; Column 29-34; 38pp; English.
XX
XX This sequence encodes the human retinoic acid receptor alpha, RAR-alpha
CC which is used in a method for identifying a subject with acute
CC promyelocytic leukaemia (APL) resulting from a t(15;17) translocation who
CC will respond to treatment with all-trans retinoic acid. The protein can
CC also be used to identify a subject with indications of APL who will not
CC respond to treatment with all-trans retinoic acid.
XX
XX Sequence 2928 BP; 582 A; 1011 C; 789 G; 546 T; 0 other;
SQ
Query Match 17.4%; Score 174; DB 20; Length 2928;
Best Local Similarity 100.0%; Pred. No. 7.1e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTTCACTGA 887
DB 732 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTTCACTGA 791
QY 888 ACTCTCCACCAAGTGCATCATTAAAGACTGTGGAGTTGCGCCAAAGCAGCTGCCCGCTTCAC 947
DB 792 ACTCTCCACCAAGTGCATCATTAAAGACTGTGGAGTTGCGCCAAAGCAGCTGCCCGCTTCAC 851
QY 948 CACCCTCACCATCGCGACAGATCACCCCTCCTCAAGGCTGCCTGGGACAT 1001
DB 852 CACCCTCACCATCGCGACAGATCACCCCTCCTCAAGGCTGCCTGGGACAT 905
```

```
RESULT 15
ID AAN90124
XX AAN90124,standard; DNA; 2940 BP.
XX
XX AAN90124;
XX
DT 01-NOV-1989 (first entry)
XX
DE DNA of clone phrAR1.
XX
KW Clone phrAR1; DNA; retinoic acid receptor; ligand
KW complexes; human.
XX
OS Homo sapiens (Human).
XX
XX Key Location/Qualifiers
FT CDS 103..1449
FT /*tag= a
XX
XX W08905355-A.
XX
XX 15-JUN-1989.
XX
XX 01-DEC-1988; 88WO-US04284.
XX
XX 02-DEC-1987; 87US-0276536.
```

```
XX (SALK ) SALK INST FOR BIOLOGICAL STUD.
XX
XX Evans RM, Giguere V, Ong ES, Segui PS;
XX
XX WPI: 1989-192701/26.
XX P-PSDB; AAP90395.
XX
XX DNA encoding retinoic acid receptor proteins
PT - used to produce proteins for studying complexes with
PT ligands and in diagnostic assays.
XX
XX Disclosure; fig. 1B; 75pp; English.
XX
XX DNA of clone phrAR1 contg. the primary sequence of a
CC protein (see corresp. AAP90395) that has ligand binding and
CC transcription activating properties of retinoic acid receptor
CC (RAR) protein. Used to make chimeric receptors,
CC to produce receptor, to study binding complexes, and to screen cpds.
CC for RAR-agonists and antagonists.
XX
XX Sequence 2940 BP; 594 A; 1010 C; 790 G; 546 T; 0 other;
SQ
Query Match 17.4%; Score 174; DB 10; Length 2940;
Best Local Similarity 100.0%; Pred. No. 7.1e-31;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTTCACTGA 887
DB 732 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGGACAAAGTTCACTGA 791
QY 888 ACTCTCCACCAAGTGCATCATTAAAGACTGTGGAGTTGCGCCAAAGCAGCTGCCCGCTTCAC 947
DB 792 ACTCTCCACCAAGTGCATCATTAAAGACTGTGGAGTTGCGCCAAAGCAGCTGCCCGCTTCAC 851
QY 948 CACCCTCACCATCGCGACAGATCACCCCTCCTCAAGGCTGCCTGGGACAT 1001
DB 852 CACCCTCACCATCGCGACAGATCACCCCTCCTCAAGGCTGCCTGGGACAT 905
```

Search completed: March 30, 2003, 10:14:03  
Job time : 214.299 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 09:12:13 ; Search time 33.7584 Seconds  
(without alignments)  
9093.534 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_10000\_11000  
Perfect score: 1001  
Sequence: 1 cccaggagactgcagctgg.....aaggctgctgcctggacat 1001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents.NA.\*  
1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	174	17.4	704	2	US-08-592-383-5
2	174	17.4	2658	2	US-08-592-383-3
3	174	17.4	2328	2	US-08-095-728B-3
4	174	17.4	2928	5	PCT-US92-02320A-3
5	174	17.4	2940	2	US-08-592-383-1
6	174	17.4	2940	6	5171671-1
7	174	17.4	3036	1	US-08-306-691B-52
8	174	17.4	3036	2	US-08-095-728B-1
9	174	17.4	3036	5	PCT-US92-02320A-1
10	174	17.4	3511	3	US-08-892-747-13
11	112.8	11.3	2989	6	5223606-1
12	109.2	10.9	1576	6	5260432-1
13	86.6	8.7	558	2	US-08-896-365-5
14	56.1	5.6	1934	4	US-08-776-844-1
15	56	5.6	1959	1	US-08-342-411A-3
16	56	5.6	1959	5	PCT-US94-12883-4
17	55	5.5	7218	1	US-08-232-463-14
18	53.8	5.4	1893	6	5438126-1
19	53	5.3	320	4	US-09-165-264-7
20	52.8	5.3	1860	2	US-08-372-652-7
21	52.8	5.3	1860	5	PCT-US95-16311-7
22	52	5.2	320	4	US-09-165-264-14
23	51.8	5.2	320	4	US-09-165-264-13
24	51.2	5.1	318	4	US-09-165-264-12
25	50.4	5.0	320	4	US-09-165-264-11
26	49.6	5.0	319	4	US-09-165-264-8
27	49.6	5.0	1688	2	US-08-649-619B-2

28	49.6	5.0	1813	5	PCT-US94-12883-3
29	49.6	5.0	1898	1	US-08-342-411A-1
30	49.6	5.0	1979	2	US-08-649-619B-1
31	49.6	5.0	2030	1	US-08-330-518-1
32	49.6	5.0	2030	1	US-08-330-283-1
33	49.6	5.0	2030	2	US-08-646-248-1
34	49.6	5.0	2030	5	PCT-US95-13924-1
35	49.6	5.0	2030	5	PCT-US95-13931-1
36	47	4.7	152331	3	US-09-128-155-16
37	45.8	4.6	816	1	US-08-485-971-22
38	45.8	4.6	816	1	US-08-383-754-22
39	45.8	4.6	816	1	US-08-485-978-22
40	45.8	4.6	816	2	US-08-486-814-22
41	45.8	4.6	816	2	US-08-487-472-22
42	45.8	4.6	816	3	US-08-485-740-22
43	45.8	4.6	816	3	US-09-162-184-22
44	45.8	4.6	816	4	US-09-161-902-22
45	45.8	4.6	816	4	US-09-489-777A-22

## ALIGNMENTS

RESULT 1  
US-08-592-383-5  
; Sequence 5, Application US/08592383  
; Patent No. 5830760  
; GENERAL INFORMATION:  
; APPLICANT: Tsai, S. and S.J. Collins  
; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acti  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage  
; COMPUTER: IBM PC/386 Compatible  
; OPERATING SYSTEM: MS-DOS 4.01  
; SOFTWARE: Word for Windows 5.01-t  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/592,383  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/099,242  
; FILING DATE: July 28, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Broderick, Thomas F.  
; REGISTRATION NUMBER: 31,332  
; REFERENCE/DOCKET NUMBER: FHCR-1-7190  
; TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 704 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: CDNA to mRNA  
; DESCRIPTION: page 11; RAR-alpha RA binding region; positions 692 to 1395  
; IMMEDIATE SOURCE:  
; LIBRARY: CDNA  
; US-08-592-383-5

Query Match 17.4%; Score 174; DB 2; Length 704;  
Best Local Similarity 100.0%; Pred. No. 4.3e-34;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACAGTGTCTCTCTGGACATTGACCTTGGGACAAAGTTCAGTGA 887

Db 41 GAACAACAGCTCAGAACACGTGTCTCTGGACATTGACCTCGGACAAAGTTTCAGTGA 100  
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGTGCCTGGCCTTCAC 947  
Db 101 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGTGCCTGGCCTTCAC 160  
QY 948 CACCTCACCATTGCCGACCAAGATCACCCTCTCAAGCTCCTGCTGGGACAT 1001  
Db 161 CACCTCACCATTGCCGACCAAGATCACCCTCTCAAGCTCCTGCTGGGACAT 214

## RESULT 2

US-08-592-383-3  
; Sequence 3, Application US/08592383  
; Patent No. 5830760  
; GENERAL INFORMATION:  
; APPLICANT: Tsai, S. and S.J. Collins  
; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acid Recept  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage  
; COMPUTER: IBM PC/386 Compatible  
; OPERATING SYSTEM: MS-DOS 4.01  
; SOFTWARE: Word for Windows 5.01-t  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/592,383  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/099,242  
; FILING DATE: July 28, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Broderick, Thomas F.  
; REGISTRATION NUMBER: 31,332  
; REFERENCE/DOCKET NUMBER: FHCR-1-7190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2658 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; DESCRIPTION: page 4, RAR-alpha403 dominant negative; deleted of 1311-1596 of SEQ  
; IMMEDIATE SOURCE:  
; LIBRARY: cDNA  
US-08-592-383-3

Query Match 17.4%; Score 174; DB 2; Length 2658;  
Best Local Similarity 100.0%; Pred. No. 6.6e-34;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTGTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 887  
Db 732 GAACAACAGCTCAGAACACGTGTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 791  
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGTGCCTGGCCTTCAC 947  
Db 792 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGTGCCTGGCCTTCAC 851  
QY 948 CACCTCACCATTGCCGACCAAGATCACCCTCTCAAGCTCCTGCTGGGACAT 1001  
Db 852 CACCTCACCATTGCCGACCAAGATCACCCTCTCAAGCTCCTGCTGGGACAT 905

## RESULT 3

US-08-095-728B-3  
; Sequence 3, Application US/08095728B  
; Patent No. 5843642  
; GENERAL INFORMATION:  
; APPLICANT: DMITROVSKY, ETHAN  
; APPLICANT: WARRELL JR, RAYMOND P  
; APPLICANT: MILLER JR, WILSON H  
; APPLICANT: FRANKEL, STANLEY  
; TITLE OF INVENTION: METHODS FOR THE DETECTION AND  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOPER & DUNHAM LLP  
; STREET: 1185 AVENUE OF THE AMERICAS  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/095,728B  
; FILING DATE: 21-JUL-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/673,838  
; FILING DATE: 22-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WHITE, JOHN P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 38694-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2928 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; IMMEDIATE SOURCE:  
; CLONE: hRAR ALPHA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 103..1488  
; OTHER INFORMATION:  
US-08-095-728B-3

Query Match 17.4%; Score 174; DB 2; Length 2928;  
Best Local Similarity 100.0%; Pred. No. 6.8e-34;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTGTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 887  
Db 732 GAACAACAGCTCAGAACACGTGTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 791  
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGTGCCTGGCCTTCAC 947  
Db 792 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCCAAGCAGTGCCTGGCCTTCAC 851  
QY 948 CACCTCACCATTGCCGACCAAGATCACCCTCTCAAGCTCCTGCTGGGACAT 1001  
Db 852 CACCTCACCATTGCCGACCAAGATCACCCTCTCAAGCTCCTGCTGGGACAT 905

RESULT 4  
PCT-US92-02320A-3

Sequence 3, Application PC/TUS9202320A  
GENERAL INFORMATION:  
APPLICANT: Sloan-Kettering Institute, For Cancer Research  
TITLE OF INVENTION: METHODS FOR DETECTION AND TREATMENT OF CANCER  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: COOPER & DUNHAM  
STREET: 30 ROCKEFELLER PLAZA  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/02320A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 675,084  
FILING DATE: 22-MAR-1991  
PRIOR APPLICATION DATA: US 673,838  
FILING DATE: 22-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: WHITE, JOHN P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 38694-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 644-0525  
TELEX: (212) 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2928 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cdna  
IMMEDIATE SOURCE:  
CLONE: hRAR ALPHA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 103..1488  
OTHER INFORMATION:  
PCT-US92-02320A-3  
Query Match 17.4%; Score 174; DB 5; Length 2928;  
Best Local Similarity 100.0%; Pred. No. 6.8e-34;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 828 GAACAACAGCTCAGAACAGCTGCTCTCTGGACATTGACCTCTGGGACAGTTCAGTGA 887  
Db 732 GAACAACAGCTCAGAACAGCTGCTCTCTGGACATTGACCTCTGGGACAGTTCAGTGA 791  
QY 888 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 947  
Db 792 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 851  
QY 948 CACCCTCACCATCGCGGACAGATCACCCCTCTCAAGGCTGCCTGCCTGGACAT 1001  
Db 852 CACCCTCACCATCGCGGACAGATCACCCCTCTCAAGGCTGCCTGCCTGGACAT 905  
RESULT 5  
US-08-592-383-1  
Sequence 1, Application US/08592383  
Patent No. 5830760  
GENERAL INFORMATION:  
APPLICANT: Tsai, S. and S.J. Collins

TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Aci  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
STREET: 2800 Pacific First Centre, 1420 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage  
COMPUTER: IBM PC/386 Compatible  
OPERATING SYSTEM: MS-DOS 4.01  
SOFTWARE: Word for Windows 5.01-t  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,383  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/099,242  
FILING DATE: July 28, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Broderick, Thomas F.  
REGISTRATION NUMBER: 31,332  
REFERENCE/DOCKET NUMBER: FPCR-1-7190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2940 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cdna to mRNA  
DESCRIPTION: see Figure 16B; RAR-alpha  
IMMEDIATE SOURCE:  
LIBRARY: CDNA  
US-08-592-383-1  
Query Match 17.4%; Score 174; DB 2; Length 2940;  
Best Local Similarity 100.0%; Pred. No. 6.8e-34;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 828 GAACAACAGCTCAGAACAGCTGCTCTCTGGACATTGACCTCTGGGACAGTTCAGTGA 887  
Db 732 GAACAACAGCTCAGAACAGCTGCTCTCTGGACATTGACCTCTGGGACAGTTCAGTGA 791  
QY 888 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 947  
Db 792 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTCAC 851  
QY 948 CACCCTCACCATCGCGGACAGATCACCCCTCTCAAGGCTGCCTGCCTGGACAT 1001  
Db 852 CACCCTCACCATCGCGGACAGATCACCCCTCTCAAGGCTGCCTGCCTGGACAT 905  
RESULT 6  
5171671-1  
Patent No. 5171671  
APPLICANT: EVANS, RONALD M.; ONG, ESTELITA S.; SEGUI,  
PRUDHAR S.; THOMPSON, CATHERINE C.; UEMONO, KAZUHIKO  
GUGUERE, VINCENT  
TITLE OF INVENTION: RETINOIC ACID RECEPTOR COMPOSITION  
NUMBER OF SEQUENCES: 2  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/546,256  
FILING DATE: 06-AUG-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 276,536  
FILING DATE: 30-NOV-1988  
APPLICATION NUMBER: 128,331  
FILING DATE: 02-DEC-1987  
SEQ ID NO:1;

LENGTH: 2940  
5171671-1

Query Match 17.4%; Score 174; DB 6; Length 2940;  
Best Local Similarity 100.0%; Pred. No. 6.8e-34;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTTCAGTGA 887  
|||||  
Db 732 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTTCAGTGA 791  
|||||  
QY 888 ACTCTCCACCAAGTCATCATTAAAGACTGTGGAGTTGCCCAAGCAGCTGCCCGGCTTCAC 947  
|||||  
Db 792 ACTCTCCACCAAGTCATCATTAAAGACTGTGGAGTTGCCCAAGCAGCTGCCCGGCTTCAC 851  
|||||  
QY 948 CACCTCACCATCGCGGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 1001  
|||||  
Db 852 CACCTCACCATCGCGGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 905  
|||||

## RESULT 7

US-08-306-691B-52  
; Sequence 52, Application US/08306691B  
; Patent No. 5734039  
; GENERAL INFORMATION:  
; APPLICANT: Calabretta, Bruno  
; APPLICANT: Skorski, Tomasz  
; TITLE OF INVENTION: ANTISENSE  
; TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES  
; NUMBER OF SEQUENCES: 55  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Seidel, Gonda, Lavorigna & Monaco, P.C.  
; STREET: Two Penn Center, Suite 1800  
; CITY: Philadelphia  
; STATE: Pennsylvania  
; COUNTRY: U.S.A.  
; ZIP: 19102

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 720 Kb  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/306,691B  
FILING DATE: September 15, 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Monaco, Daniel A.  
REGISTRATION NUMBER: 30,480  
REFERENCE/DOCKET NUMBER: 8321-8  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-8383  
TELEFAX: (215) 568-5549  
TELEX: No. 5734039e  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3036 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear

US-08-306-691B-52  
Query Match 17.4%; Score 174; DB 1; Length 3036;  
Best Local Similarity 100.0%; Pred. No. 6.9e-34;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTTCAGTGA 887  
|||||  
Db 1701 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTTCAGTGA 1760  
|||||

QY 888 ACTCTCCACCAAGTCATCATTAAAGACTGTGGAGTTGCCCAAGCAGCTGCCCGGCTTCAC 947  
|||||  
Db 1761 ACTCTCCACCAAGTCATCATTAAAGACTGTGGAGTTGCCCAAGCAGCTGCCCGGCTTCAC 1820  
|||||  
QY 948 CACCTCACCATCGCGGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 1001  
|||||  
Db 1821 CACCTCACCATCGCGGACGACATCACCTCTCAAGGCTGCTGCTGGACAT 1874  
|||||

## RESULT 8

US-08-095-728B-1  
; Sequence 1, Application US/08095728B  
; Patent No. 5843642  
; GENERAL INFORMATION:  
; APPLICANT: DMITROVSKY, ETHAN  
; APPLICANT: WARRELL JR, RAYMOND P  
; APPLICANT: MILLER JR, WILSON H  
; APPLICANT: FRANKEL, STANLEY  
; TITLE OF INVENTION: METHODS FOR THE DETECTION AND  
; TITLE OF INVENTION: TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA (APL)  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOPER & DUNHAM LLP  
; STREET: 1185 AVENUE OF THE AMERICAS  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10036

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/095,728B  
FILING DATE: 21-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/673,838  
FILING DATE: 22-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: WHITE, JOHN P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 38694-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3036 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
CLONE: MYL-RAR  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 67..2457  
OTHER INFORMATION:

US-08-095-728B-1

Query Match 17.4%; Score 174; DB 2; Length 3036;  
Best Local Similarity 100.0%; Pred. No. 6.9e-34;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTTCAGTGA 887  
|||||  
Db 1701 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTTCAGTGA 1760  
|||||

QY 888 ACTCTCCACCAAGTCATCATTAAAGACTGTGGAGTTGCCCAAGCAGCTGCCCGGCTTCAC 947  
|||||  
Db 1761 ACTCTCCACCAAGTCATCATTAAAGACTGTGGAGTTGCCCAAGCAGCTGCCCGGCTTCAC 1820  
|||||

QY 948 CACCCTCACCATCGCGGACAGATCACCCCTCTCAAGGCTGCCTGGACAT 1001  
|||||  
Db 1821 CACCCTCACCATCGCGGACAGATCACCCCTCTCAAGGCTGCCTGGACAT 1874  
|||||

## RESULT 9

PCT-US92-02320A-1

; Sequence 1, Application PC/TUS9202320A  
; GENERAL INFORMATION:  
; APPLICANT: Sloan-Kettering Institute, For Cancer Research  
; TITLE OF INVENTION: METHODS FOR DETECTION AND TREATMENT OF CANCER  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOPER & DUNHAM  
; STREET: 30 ROCKEFELLER PLAZA  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/02320A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 675,084  
; FILING DATE: 22-MAR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 673,838  
; FILING DATE: 22-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WHITE, JOHN P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 38694-PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 644-0525  
; TELEX: (212) 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3036 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; IMMEDIATE SOURCE:  
; CLONE: MYL-RAR  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 67..2457  
; OTHER INFORMATION:  
; PCT-US92-02320A-1

Query Match 17.4%; Score 174; DB 5; Length 3036;  
Best Local Similarity 100.0%; Pred. No. 6.9e-34;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTTCAGTGA 887  
|||||  
Db 1701 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTTCAGTGA 1760  
|||||  
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTTCGCCAAGCAGTGCCCGGCTTCAC 947  
|||||  
Db 1761 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTTCGCCAAGCAGTGCCCGGCTTCAC 1820  
|||||  
QY 948 CACCCTCACCATCGCGGACAGATCACCCCTCTCAAGGCTGCCTGGACAT 1001  
|||||  
Db 1821 CACCCTCACCATCGCGGACAGATCACCCCTCTCAAGGCTGCCTGGACAT 1874  
|||||

## RESULT 10

US-08-892-747-13  
; Sequence 13, Application US/08892747  
; Patent No. 6057153  
; GENERAL INFORMATION:  
; APPLICANT: Shaji T. George, Michael Ma, Martina Werner,  
; APPLICANT: Umberto Pace and Allan R. Goldberg  
; TITLE OF INVENTION: Stabilized External Guide Sequences  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/892,747  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/372,556  
; FILING DATE: January 13, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/00513  
; FILING DATE: January 19, 1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
; REGISTRATION NUMBER: 31,284  
; REFERENCE/DOCKET NUMBER: ILI109CIP2  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (404) 873-8794  
; TELEFAX: (404) 873-8795  
; INFORMATION FOR SEQ ID NO: 13:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3511 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: 1..3511  
; OTHER INFORMATION: /function= "PML-RAR" DNA Sequence."  
; US-08-892-747-13

Query Match 17.4%; Score 174; DB 3; Length 3511;  
Best Local Similarity 100.0%; Pred. No. 7.2e-34;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTTCAGTGA 887  
|||||  
Db 2176 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGACAAAGTTTCAGTGA 2235  
|||||  
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTTCGCCAAGCAGTGCCCGGCTTCAC 947  
|||||  
Db 2236 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTTCGCCAAGCAGTGCCCGGCTTCAC 2295  
|||||  
QY 948 CACCCTCACCATCGCGGACAGATCACCCCTCTCAAGGCTGCCTGGACAT 1001  
|||||  
Db 2296 CACCCTCACCATCGCGGACAGATCACCCCTCTCAAGGCTGCCTGGACAT 2349  
|||||

APPLICANT: Messer, Lori A.  
APPLICANT: Tun-ping, Yu  
TITLE OF INVENTION: GENES AND GENETIC MARKERS FOR IMPROVED  
TITLE OF INVENTION: REPRODUCTIVE TRAITS IN ANIMALS  
NUMBER OF SEQUENCES: 25  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Zarley, McKee, Thomte, Voorhees & Sease  
STREET: 801 Grand Avenue, Suite 3200  
CITY: Des Moines  
STATE: Iowa  
COUNTRY: USA  
ZIP: 50309  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

Query Match	8.7%	Score 86.6;	DB 2;	Length 558;
Best Local Similarity	74.2%	Pred. No. 1.4e-13;		
Matches 135;	Conservative	0;	Mismatches 45;	Indels 2;
	Gaps			2;

RESULT 14  
US-08-776-844-1  
; Sequence 1, Application US/08776844  
; Patent No. 6277976  
; GENERAL INFORMATION:  
; APPLICANT: ENMARK, EVA

	Query Match	11.3%	Score 112.8	DB 6	Length 2989
	Best Local Similarity	76.7%	Pred. No. 8.6e-19		
	Matches 135	Conservative 0	Mismatches 42	Indels 0	Gaps 0
QY	822	CCCCCAGAACAAACAGCTCAGAACAAACGTGTCTCTCTGTGACATTCACCTCTCGGACAAAGTT	881		
DB	924	CACCACCAAAATCCAGTGTGACCATCGAGTCGCGACTGGACCTGGCGCTCTCGGACAAAT	983		
QY	882	CAGTGAAGACTCTCCACCAAGTGCATTAAGACATGTGGAGTTGCGCAAGCAGCTGCCCGG	941		
DB	984	CAGTGAAGTGGCCACCAAGTGCATTTAAGATCGTGGAGTTTGCTAAACGCTCTGCCTGG	1043		
QY	942	CTTCACCAACCGTCAACATATGCGCGACCAAGATCACCCCTCTCTCAAGGTCGCTCGCTGGACAT	1001		
DB	1044	TTTTCACGTGGGTGACCATCGAGACCAAAATPACCGTCTGTAAGGCGCGCTGCCCTGGAGAT	1103		

RESULT 12.  
5260432-1  
; Patent No. 5260432  
; APPLICANT: TAKAKU, FUMIMARO; ISHIKAWA, TAKAKASHI; IMAWARI, MICHIO;  
; EVANS, RONALD. M.; UMESONO, KAZUHIKO  
; TITLE OF INVENTION: HUMAN GAMMA RETINOIC ACID RECEPTOR DNA  
; NUMBER OF SEQUENCES: 4  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/370,407  
; FILING DATE: 22-JUN-1989  
; SEQ ID NO:1: ;  
; LENGTH: 1576  
5260432-1

	Query Match	10.9%;	Score 109.2;	DB 6;	Length 1576;
	Best Local Similarity	75.8%;	Pred. No. 5.4e-18;		
	Matches 135;	Conservative	0;	Mismatches 43;	Indels 0; Gaps 0;
QY	824	CCGAGACACAGCTCAGACAAACGTCGCTCTCTCGGACATTTGACCTCTGGGACACAAGTTCA	883		
Db	831	CCAGGAACCTCAAGTCGACAGACACCGCTGCGAGCTGGATCTGGGGCTCTGGGACAAAGTTCA	890		
QY	884	GTGAACCTCTCCACCAAGTGCATCATTAAGACGTGTGGAGTTGCCCAAGCAGCTGCCCGGCT	943		
Db	891	GTGAGCTGGCTACCAAGTGATCATCATGATCGTGGAGTTTGCCCAAGCGGTTGGCTGGCT	950		
QY	944	TCACCAACCTTCACCATCGCCGACCCAGATFACCCCTCTCAAGGCTGCCCTGGCTGGACAT	1001		
Db	951	TTACAGGGCTCAGCATTTGCTCACCAGATCATCTCTGCTCAAAAGCTGCCCTGCCTAGATAT	1008		

RESULT 13  
US-08-896-365-5  
; Sequence 5, Application US/08896365  
; Patent No. 5939264  
; GENERAL INFORMATION:  
; APPLICANT: Rothschild, Max F.  
; APPLICANT: Tuggle, Christopher K



APPLICANT: GUSTAFSSON, JAN  
TITLE OF INVENTION: OR-1 ON ORPHAN RECEPTOR BELONGING  
TO THE NUCLEAR RECEPTOR FAMILY  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Banner & Witcoff  
STREET: 1001 G Street, NW  
CITY: Washington  
STATE: DC  
COUNTRY: USA  
ZIP: 20001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/776,844  
FILING DATE: 24-JUN-1997  
CLASSIFICATION: 536  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/03247  
FILING DATE: 16-AUG-1995  
APPLICATION NUMBER: UK 9413536.2  
FILING DATE: 16-AUG-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Kagan, Sarah A  
REGISTRATION NUMBER: 32141  
REFERENCE/DOCKET NUMBER: 00487.04029  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-508-9100  
TELEFAX: 202-508-9299  
TELEX:  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1934 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
US-08-776-844-1

Query Match 5.6%; Score 56; DB 4; Length 1934;  
Best Local Similarity 63.2%; Pred. No. 7.7e-05;  
Matches 86; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 866 ACCTCTGGGACAGTTTCAGTGAACCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGG 925  
Db 942 AACGCTTTGCCCACTTCACTAGCTAGCCATCATCTCAGTCCAGGAGATCGTGGACTTCG 1001  
QY 926 CCAAGCAGGTGCCCGCTTCCACCCCTCACCATCGCCGACCATCACCCTCTCTCAAGG 985  
Db 1002 CCAAGCAGGTGCCCGCTTCCAGCTGGCGGGAGGACCATCAGATCGCCCTCTCTGAAGG 1061  
QY 986 CTGCGCTGCTGGACAT 1001  
Db 1062 CATCCACCATCGAGAT 1077

## RESULT 15

US-08-342-411A-3  
Sequence 3, Application US/08342411A  
Patent No. 5639616  
GENERAL INFORMATION:  
APPLICANT: LIAO, Shutsung  
APPLICANT: SONG, Ching  
TITLE OF INVENTION: UBIQUITOUS NUCLEAR RECEPTOR:  
COMPOSITIONS AND METHODS  
NUMBER OF SEQUENCES: 38  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Arnold, White & Durkee  
STREET: P.O. Box 4433  
CITY: Houston

STATE: TX  
COUNTRY: USA  
ZIP: 77210-4433  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC Compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/342,411A  
FILING DATE: 18-NOV-1994  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: KITCHELL, BARBARA S.  
REGISTRATION NUMBER: 33,928  
REFERENCE/DOCKET NUMBER: ARCD154  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (512) 418-3000  
TELEFAX: (713) 789-2679  
TELEX: 79-0924  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1959 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 256..1584  
US-08-342-411A-3

Query Match 5.6%; Score 56; DB 1; Length 1959;  
Best Local Similarity 63.2%; Pred. No. 7.8e-05;  
Matches 86; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 866 ACCTCTGGGACAGTTTCAGTGAACCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGG 925  
Db 998 AACGCTTTGCCCACTTCACTAGCTAGCCATCATCTCAGTCCAGGAGATCGTGGACTTCG 1057  
QY 926 CCAAGCAGCTGCCCGCTTCCACCCCTCACCATCGCCGACCATCACCCTCTCTCAAGG 985  
Db 1058 CCAAGCAGTGCAGGGTTCCTGCGAGCTGGCGGGAGGACCATGTCCTCTCTGAAGG 1117  
QY 986 CTGCGCTGCTGGACAT 1001  
Db 1118 CATCCACCATCGAGAT 1133  
Search completed: March 30, 2003, 13:56:34  
Job time : 45.7584 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 13:12:08 ; Search time 75.7689 Seconds  
(without alignments)  
11242.392 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_10000\_11000  
Perfect score: 1001  
Sequence: 1 cccaggagactgcagctgg.....aaggctgctgcctgacat 1001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 574371 seqs, 425486471 residues

Total number of hits satisfying chosen parameters: 1148742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications\_NA.\*

- 1: /cgn2.6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2.6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2.6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2.6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2.6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2.6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2.6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2.6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2.6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2.6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2.6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2.6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2.6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2.6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match	Description
1	174	17.4	Sequence 524, Appl
2	166	16.6	Sequence 4521, Ap
3	166	16.6	Sequence 5714, Ap
4	166	16.6	Sequence 2138, Ap
5	166	16.6	Sequence 12806, A
6	166	16.6	Sequence 31, Appl
7	117.6	11.7	Sequence 7, Appl
8	117.6	11.7	Sequence 8, Appl
9	117.6	11.7	Sequence 9, Appl
10	117.6	11.7	Sequence 10, Appl
11	109.2	10.9	Sequence 1, Appl
12	99.6	10.0	Sequence 1210, Ap
13	89	8.9	Sequence 92, Appl
14	56	5.6	Sequence 33, Appl
15	56	5.6	Sequence 1, Appl
16	56	5.6	Sequence 1, Appl
17	56	5.6	Sequence 1, Appl
18	53.8	5.4	Sequence 567, App
19	52.8	5.3	Sequence 1, Appl

20	51.8	5.2	12733	9	US-10-032-393-47	Sequence 47, Appl
21	51.8	5.2	12739	9	US-10-032-393-8	Sequence 8, Appl
22	49.6	5.0	794	9	US-10-043-487-53	Sequence 53, Appl
23	49.6	5.0	1622	10	US-09-925-297-176	Sequence 176, Appl
24	47.2	4.7	1064	10	US-09-804-682-29	Sequence 29, Appl
25	47	4.7	152331	9	US-10-095-407-16	Sequence 16, Appl
26	46.4	4.6	1390	10	US-09-760-364-8	Sequence 8, Appl
27	45.8	4.6	816	9	US-09-921-650-22	Sequence 22, Appl
28	45.8	4.6	1416	10	US-09-874-389-22	Sequence 22, Appl
29	45.8	4.6	1443	9	US-09-921-650-37	Sequence 37, Appl
30	44.6	4.5	440	9	US-10-184-644-202	Sequence 202, Appl
31	44.6	4.5	1332	10	US-09-883-093-1	Sequence 1, Appl
32	44.4	4.4	987	10	US-09-804-682-20	Sequence 20, Appl
33	44.2	4.4	428	10	US-09-864-761-23488	Sequence 23488, A
34	44.2	4.4	577	10	US-09-864-761-6755	Sequence 6755, Ap
35	44.2	4.4	693	12	US-10-044-090-315	Sequence 315, App
36	44.2	4.4	1528	10	US-09-962-832-109	Sequence 109, App
37	44.2	4.4	1528	10	US-09-880-107-3328	Sequence 3328, Ap
38	44.2	4.4	1679	12	US-10-044-090-316	Sequence 316, Appl
39	44	4.4	197496	9	US-09-877-177-10	Sequence 10, Appl
40	43.6	4.4	450	10	US-09-960-352-10528	Sequence 10528, A
41	42.2	4.2	2581	10	US-09-742-732-1	Sequence 1, Appl
42	42	4.2	405	10	US-09-960-352-13861	Sequence 13861, A
43	42	4.2	1450	10	US-09-880-107-3875	Sequence 3875, Ap
44	41.6	4.2	451	10	US-09-864-761-19039	Sequence 19039, A
45	41.6	4.2	458	10	US-09-864-761-2304	Sequence 2304, Ap

## ALIGNMENTS

RESULT 1  
US-09-954-456-524  
; Sequence 524, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Us:  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; CURRENT FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 524  
; LENGTH: 2907  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-456-524

Query Match 17.4%; Score 174; DB 10; Length 2907;  
Best Local Similarity 100.0%; Pred. No. 1.2e-38;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGCAAGTTCAGTGA 887  
DB 732 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGCAAGTTCAGTGA 791  
QY 888 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 947  
DB 792 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 851  
QY 948 CACCTCCACCATCGCGGACAGATCACCTCTCAAGGCTGCTCGCTGGACAT 1001  
DB 852 CACCTCCACCATCGCGGACAGATCACCTCTCAAGGCTGCTCGCTGGACAT 905

RESULT 2  
US-09-960-352-4521  
; Sequence 4521, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 4521  
; LENGTH: 393  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 20-LIB188-012-Q1-E1-E11  
US-09-960-352-4521

Query Match 16.6%; Score 166; DB 10; Length 393;  
Best Local Similarity 97.1%; Pred. No. 1.4e-36;  
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGCAAGTTCAGTGA 887  
DB 206 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGCAAGTTCAGTGA 265  
QY 888 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 947  
DB 266 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 325  
QY 948 CACCTCCACCATCGCGGACAGATCACCTCTCAAGGCTGCTCGCTGGACAT 1001  
DB 326 CACCTCCACCATCGCGGACAGATCACCTCTCAAGGCTGCTCGCTGGATAT 379

RESULT 3  
US-09-960-352-5714  
; Sequence 5714, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 5714  
; LENGTH: 411  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; OTHER INFORMATION: Clone ID: 25-LIB188-001-Q1-E1-G1

US-09-960-352-5714

Query Match 16.6%; Score 166; DB 10; Length 411;  
Best Local Similarity 97.1%; Pred. No. 1.4e-36;  
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGCAAGTTCAGTGA 887  
DB 206 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGCAAGTTCAGTGA 265  
QY 888 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 947  
DB 266 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 325  
QY 948 CACCTCCACCATCGCGGACAGATCACCTCTCAAGGCTGCTCGCTGGACAT 1001  
DB 326 CACCTCCACCATCGCGGACAGATCACCTCTCAAGGCTGCTCGCTGGATAT 379

RESULT 4  
US-09-960-352-2138  
; Sequence 2138, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 2138  
; LENGTH: 416  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (76)  
; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: 10-LIB188-020-Q1-E1-C5  
US-09-960-352-2138

Query Match 16.6%; Score 166; DB 10; Length 416;  
Best Local Similarity 97.1%; Pred. No. 1.4e-36;  
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;  
QY 828 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGCAAGTTCAGTGA 887  
DB 142 GAACAACAGCTCAGAACACGCTGCTCTCTGGACATTGACCTCTGGCAAGTTCAGTGA 201  
QY 888 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 947  
DB 202 ACTCTCCACCAGTGCATCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 261  
QY 948 CACCTCCACCATCGCGGACAGATCACCTCTCAAGGCTGCTCGCTGGACAT 1001  
DB 262 CACCTCCACCATCGCGGACAGATCACCTCTCAAGGCTGCTCGCTGGATAT 315

RESULT 5  
US-09-960-352-12806/c  
; Sequence 12806, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathialagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION

FILE REFERENCE: 16511.006/37-21(10298)C  
CURRENT APPLICATION NUMBER: US/09/960,352  
CURRENT FILING DATE: 2001-09-24  
NUMBER OF SEQ ID NOS: 15112  
SEQ ID NO 12806  
LENGTH: 417  
TYPE: DNA  
ORGANISM: Bos taurus  
OTHER INFORMATION: Clone ID: 55-LIB188-005-Q1-E1-F4  
US-09-960-352-12806

Query Match : 16.6%; Score 166; DB 10; Length 417;  
Best Local Similarity 97.1%; Pred. No. 1.4e-36;  
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 887  
|||||  
DB 212 GAACAACAGCTCAGAACACGGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 153  
|||||  
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 947  
|||||  
DB 152 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 93  
|||||  
QY 948 CACCCTCACCATCGCGGACGACATCACCTCTCAAGGCTGCCTGCCTGCACAT 1001  
|||||  
DB 92 CACCCTCACCATCGCGGACGACATCACCTCTCAAGGCTGCCTGCCTGCATAT 39  
|||||

## RESULT 6

US-09-960-352-31

Sequence 31, Application US/09960352

Patent No. US20020137139A1

GENERAL INFORMATION:

APPLICANT: Warrent, Wesley C.

APPLICANT: Tao, Nengping

APPLICANT: Byatt, John C.

APPLICANT: Mathialagan, Nagappan

TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND

TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION

FILE REFERENCE: 16511.006/37-21(10298)C

CURRENT APPLICATION NUMBER: US/09/960,352

CURRENT FILING DATE: 2001-09-24

NUMBER OF SEQ ID NOS: 15112

SEQ ID NO 31

LENGTH: 429

TYPE: DNA

ORGANISM: Bos taurus

OTHER INFORMATION: Clone ID: 01-LIB188-009-Q1-E1-A1

US-09-960-352-31

Query Match : 16.6%; Score 166; DB 10; Length 429;  
Best Local Similarity 97.1%; Pred. No. 1.4e-36;  
Matches 169; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 887  
|||||  
DB 234 GAACAACAGCTCAGAACACGGTCTCTCTGGACATTGACCTCTGGGACAAAGTTCAAGTGA 293  
|||||  
QY 888 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 947  
|||||  
DB 294 ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGGCTTCAC 353  
|||||  
QY 948 CACCCTCACCATCGCGGACGACATCACCTCTCAAGGCTGCCTGCCTGCACAT 1001  
|||||  
DB 354 CACCCTCACCATCGCGGACGACATCACCTCTCAAGGCTGCCTGCCTGCATAT 407  
|||||

## RESULT 7

US-10-239-804-7

Sequence 7, Application US/10239804

Publication No. US20030053991A1

GENERAL INFORMATION:

APPLICANT: Oxford Biomedica (UK) Limited

APPLICANT: Kingsman, Alan J

APPLICANT: Maden, Malcolm

APPLICANT: Corcoran, Jonathan PT

TITLE OF INVENTION: Factor

FILE REFERENCE: P009156WOCTH

CURRENT APPLICATION NUMBER: US/10/239,804

CURRENT FILING DATE: 2002-09-23

PRIOR APPLICATION NUMBER: PCT/GB00/01211

PRIOR FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: GB 0024300.6

PRIOR FILING DATE: 2000-10-04

NUMBER OF SEQ ID NOS: 73

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 8

LENGTH: 1399

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: FLAG RARbeta2

OTHER INFORMATION: PCR product

US-10-239-804-8

Query Match : 11.7%; Score 117.6; DB 9; Length 1399;  
Best Local Similarity 78.3%; Pred. No. 4e-23;  
Matches 141; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 822 CCCCAGAACACAGCTCAGAACACGTTCTCTCTGACATTGACCTCTGGGACAAAGTT 881  
|||||

APPLICANT: Kingsman, Alan J  
APPLICANT: Maden, Malcolm  
APPLICANT: Corcoran, Jonathan PT  
TITLE OF INVENTION: Factor  
FILE REFERENCE: P009156WOCTH  
CURRENT APPLICATION NUMBER: US/10/239,804  
CURRENT FILING DATE: 2002-09-23  
PRIOR APPLICATION NUMBER: PCT/GB00/01211  
PRIOR FILING DATE: 2000-03-30  
PRIOR APPLICATION NUMBER: GB 0024300.6  
PRIOR FILING DATE: 2000-10-04  
NUMBER OF SEQ ID NOS: 73  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 7  
LENGTH: 1375  
TYPE: DNA  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: RARbeta2 PCR  
OTHER INFORMATION: product  
US-10-239-804-7

Query Match : 11.7%; Score 117.6; DB 9; Length 1375;  
Best Local Similarity 78.3%; Pred. No. 3.9e-23;  
Matches 141; Conservative 0; Mismatches 39; Indels 0; Gaps 0;

QY 822 CCCCAGAACACAGCTCAGAACACGTTCTCTCTGACATTGACCTCTGGGACAAAGTT 881  
|||||  
DB 619 CACCACGAATTCACAGCGTGACACCGGTCGATGACCTCTGGGACAAAT 678  
|||||  
QY 882 CAGTGAAGTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGG 941  
|||||  
DB 679 CAGTGAAGTCTCCACCAAGTGCATTAAGACTGTGGAGTTGCGCAAGCAGCTGCCCGG 738  
|||||  
QY 942 CTTCCACCACTCACCACGCGGACGACATCACCTCTCAAGGCTGCCTGCCTGCACAT 1001  
|||||  
DB 739 CTTCCACAGGCTGACCATCGGACGACATCACCTCTCAAGGCGGCTGCTTGATAT 798  
|||||

## RESULT 8

US-10-239-804-8

Sequence 8, Application US/10239804

Publication No. US20030053991A1

GENERAL INFORMATION:

APPLICANT: Oxford Biomedica (UK) Limited

APPLICANT: Kingsman, Alan J

APPLICANT: Maden, Malcolm

APPLICANT: Corcoran, Jonathan PT

TITLE OF INVENTION: Factor

FILE REFERENCE: P009156WOCTH

CURRENT APPLICATION NUMBER: US/10/239,804

CURRENT FILING DATE: 2002-09-23

PRIOR APPLICATION NUMBER: PCT/GB00/01211

PRIOR FILING DATE: 2000-03-30

PRIOR APPLICATION NUMBER: GB 0024300.6

PRIOR FILING DATE: 2000-10-04

NUMBER OF SEQ ID NOS: 73

SOFTWARE: PatentIn Ver. 2.1

SEQ ID NO 8

LENGTH: 1399

TYPE: DNA

ORGANISM: Artificial Sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: FLAG RARbeta2

OTHER INFORMATION: PCR product

US-10-239-804-8

Db 643 CACCACGAATTCAGCGCTGACCACCGGTCGAGATTGGCTCTGGGACAAATT 702  
QY 882 CAGTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 941  
Db 703 CAGTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 762  
QY 942 CTTACACAGCTCTGACCATCGCAGACGACATCAACCTCTCCCAAGGCTGCCTGGCAGCAT 1001  
Db 763 CTTACAGGCTCTGACCATCGCAGACGACATCAACCTCTCCCAAGGCTGCCTGGCAGCAT 822

RESULT 9  
US-10-239-804-9  
; Sequence 9, Application US/10239804  
; Publication No. US20030053991A1  
; GENERAL INFORMATION:  
; APPLICANT: Oxford Biomedica (UK) Limited  
; APPLICANT: Kingsman, Alan J  
; APPLICANT: Maden, Malcolm  
; APPLICANT: Corcoran, Jonathan PT  
; TITLE OF INVENTION: Factor  
; FILE REFERENCE: P009156WOCTH  
; CURRENT APPLICATION NUMBER: US/10/239,804  
; CURRENT FILING DATE: 2002-09-23  
; PRIOR APPLICATION NUMBER: PCT/GB00/01211  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR APPLICATION NUMBER: GB 0024300.6  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 9  
; LENGTH: 9127  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: PONY-RARBeta2  
; OTHER INFORMATION: vector genome plasmid  
US-10-239-804-9

Query Match 11.7%; Score 117.6; DB 9; Length 9127;  
Best Local Similarity 78.3%; Pred. No. 5.3e-23;  
Matches 141; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
QY 822 CCCCAGAACACAGCTCAGAACAGCTGTCTCTCTGGACATTCAGCTCTGGGACAAATT 881  
Db 343 CACCACGAATTCAGCGCTGACCCGCGGTCCGATTGGAGTTCGGGCTCTGGGACAAATT 3502  
QY 882 CAGTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 941  
Db 3503 CAGTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 3562  
QY 942 CTTACACAGCTCTGACCATCGCAGACGACATCAACCTCTCCCAAGGCTGCCTGGCAGCAT 1001  
Db 3563 CTTACAGGCTCTGACCATCGCAGACGACATCAACCTCTCCCAAGGCTGCCTGGCAGCAT 3622

RESULT 10  
US-10-239-804-10  
; Sequence 10, Application US/10239804  
; Publication No. US20030053991A1  
; GENERAL INFORMATION:  
; APPLICANT: Oxford Biomedica (UK) Limited  
; APPLICANT: Kingsman, Alan J  
; APPLICANT: Maden, Malcolm  
; APPLICANT: Corcoran, Jonathan PT  
; TITLE OF INVENTION: Factor  
; FILE REFERENCE: P009156WOCTH  
; CURRENT APPLICATION NUMBER: US/10/239,804  
; CURRENT FILING DATE: 2002-09-23  
; PRIOR APPLICATION NUMBER: PCT/GB00/01211  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR APPLICATION NUMBER: GB 0024300.6  
; PRIOR FILING DATE: 2000-10-04

; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 10  
; LENGTH: 9151  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:  
; OTHER INFORMATION: PONY-FLAG-RARBeta2 vector genome plasmid  
US-10-239-804-10

Query Match 11.7%; Score 117.6; DB 9; Length 9151;  
Best Local Similarity 78.3%; Pred. No. 5.3e-23;  
Matches 141; Conservative 0; Mismatches 39; Indels 0; Gaps 0;  
QY 822 CCCCAGAACACAGCTCAGAACAGCTGTCTCTCTGGACATTCAGCTCTGGGACAAATT 881  
Db 3467 CACCACGAATTCAGCGCTGACCCGCGGTCCGATTGGAGTTCGGGCTCTGGGACAAATT 3526  
QY 882 CAGTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 941  
Db 3527 CAGTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 3586  
QY 942 CTTACACAGCTCTGACCATCGCAGACGACATCAACCTCTCCCAAGGCTGCCTGGGACAT 1001  
Db 3587 CTTACAGGCTCTGACCATCGCAGACGACATCAACCTCTCCCAAGGCTGCCTGGGACAT 3646

RESULT 11  
US-09-797-727-1  
; Sequence 1, Application US/09797727  
; Patent No. US20020077457A1  
; GENERAL INFORMATION:  
; APPLICANT: The Salk Institute for Biological Studies  
; APPLICANT: TAKAKU, Fumimaro  
; TITLE OF INVENTION: GAMMA RETINOIC ACID RECEPTOR  
; FILE REFERENCE: SALK1150-3  
; CURRENT APPLICATION NUMBER: US/09/797,727  
; CURRENT FILING DATE: 2001-08-31  
; PRIOR APPLICATION NUMBER: US 08/486,325  
; PRIOR FILING DATE: 1995-06-07  
; PRIOR APPLICATION NUMBER: US 08/100,039  
; PRIOR FILING DATE: 1993-07-30  
; PRIOR APPLICATION NUMBER: PCT/US90/03564  
; PRIOR FILING DATE: 1990-06-22  
; PRIOR APPLICATION NUMBER: US 07/370,407  
; PRIOR FILING DATE: 1989-06-22  
; NUMBER OF SEQ ID NOS: 7  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 1  
; LENGTH: 1577  
; TYPE: DNA  
; ORGANISM: Human  
; FEATURE:  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: Human Retinoic Acid Receptor-gamma (hRAR-gamma)  
; NAME/KEY: CDS  
; LOCATION: (200)..(1576)  
US-09-797-727-1

Query Match 10.9%; Score 109.2; DB 10; Length 1577;  
Best Local Similarity 75.8%; Pred. No. 8.4e-21;  
Matches 135; Conservative 0; Mismatches 43; Indels 0; Gaps 0;  
QY 824 CCCAGAACACAGCTCAGAACAGCTGTCTCTCTGGACATTCAGCTCTGGGACAAATTCA 883  
Db 831 CCACGAAGTCCAGTGCAGACACCGCGTGCAGCTGGATCTGGGCTGTGGGACAAATTCA 890  
QY 884 GTGAAGTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCT 943  
Db 891 GTGAGTGGCTACCAAGTGCATCATCAAGATCGTGGAGTTGCCAAGCGGTTGCTGCTGCT 950  
QY 944 TCACACCCCTCACCATCGCGCAGACATCAACCTCTCTCAAGGCTGCCTGGGACAT 1001



Db 466 GAGGGGGGGGAGAGANAGGGGGGGGGAGGGGGGGGGGNTGGAANAGGAGG 407  
QY 365 AGCACCCTGCAGCTG 379  
Db 406 AGGAGNGNAGAGNG 392

## RESULT 15

US-09-909-446-1  
; Sequence 1, Application US/09909446  
; Patent NO. US20020052489A1  
; GENERAL INFORMATION:  
; APPLICANT: ENMARK, EVA  
; GUSTAFSSON, JAN  
; TITLE OF INVENTION: OR-1 ON ORPHAN RECEPTOR BELONGING  
; TO THE NUCLEAR RECEPTOR FAMILY  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/909,446  
; FILING DATE: 19-Jul-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/776,844  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: UK 9413536.2  
; FILING DATE: 16-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kagan, Sarah A  
; REGISTRATION NUMBER: 32141  
; REFERENCE/DOCKET NUMBER: 00487.04029  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1934 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; SEQUENCE DESCRIPTION: SEQ ID NO: 1:  
US-09-909-446-1

Query Match 5.68; Score 56; DB 10; Length 1934;  
Best Local Similarity 63.2%; Pred. No. 4.3e-06;  
Matches 86; Conservative 0; Mismatches 50; Indels 0; Gaps 0;  
QY 866 ACCCTCTGGGACAAGTTCAGTGAACCTCTCCACAGTGCATCTTAAGACTGTGGAGTTGG 925  
Db 942 AACGCTTTGCCCACTTCTACTAGCTAGCCATCATCTCAGTCAGGAGATCGTGGACTTCG 1001  
QY 926 CCAAGCAGCTCCCGGGTTTCACCACTCAGCATCGCCGACGAGATCACCTCTCAAGG 985  
Db 1002 CCAAGCAGGTGCCAGGGTTCTCGAGCTGGCCGGGAGGACCATCGCCCTCTCTGAAGG 1061  
QY 986 CTGCCTGCCTGGACAT 1001  
Db 1062 CATCCACCATCGAGAT 1077

Search completed: March 30, 2003, 16:46:16  
Job time : 94.7689 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:06:43 ; Search time 1290.57 Seconds  
(without alignments)

12561.622 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_10000\_11000

Perfect score: 1001

Sequence: 1 cccaggagactgcagctgg.....aaggctgctgctggacat 1001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*\*

2: em\_esthum:\*\*

3: em\_estin:\*\*

4: em\_estmu:\*\*

5: em\_estov:\*\*

6: em\_estpl:\*\*

7: em\_estro:\*\*

8: em\_hic:\*\*

9: gb\_est1:\*\*

10: gb\_est2:\*\*

11: gb\_hic:\*\*

12: gb\_est3:\*\*

13: gb\_est4:\*\*

14: gb\_est5:\*\*

15: em\_estfun:\*\*

16: em\_estom:\*\*

17: gb\_gss:\*\*

18: em\_gss\_hum:\*\*

19: em\_gss\_inv:\*\*

20: em\_gss\_pln:\*\*

21: em\_gss\_vrt:\*\*

22: em\_gss\_fun:\*\*

23: em\_gss\_mam:\*\*

24: em\_gss\_mus:\*\*

25: em\_gss\_other:\*\*

26: em\_gss\_pro:\*\*

27: em\_gss\_rod:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	362.6	36.2	631	13	BI016677 MR4-ET013
2	174	17.4	560	14	BM817785
3	174	17.4	646	14	BM790595
4	174	17.4	885	14	BQ706025
5	174	17.4	982	12	BE794377
6	174	17.4	992	9	AL556392

7	172.4	17.2	700	13	BI907041
8	170.4	17.0	173	12	BF848632
9	170	17.0	458	14	R73335
10	169	16.9	367	14	R71970
11	167.6	16.7	424	12	BF721043
12	167.6	16.7	764	9	AA790328
13	166.6	16.6	1073	13	BM544324
14	166.2	16.6	866	12	BF182871
15	164.6	16.4	861	10	BE547412
16	164.4	16.4	315	13	BM256373
17	161.4	16.1	510	14	BQ323167
18	161	16.1	287	12	BF087589
19	157.2	15.7	743	13	BI830768
20	155.2	15.5	812	13	BI914043
21	155.2	15.5	1034	12	BE871903
22	154.4	15.4	330	12	BE938020
23	152.6	15.2	426	12	BF229524
24	151.8	15.2	264	14	BQ329046
25	147.2	14.7	533	10	BE650995
26	145.8	14.6	328	12	BF333905
27	145	14.5	670	12	BG337888
28	140.8	14.1	456	10	AW015525
29	135.4	13.5	298	12	BF822393
30	131	13.1	270	10	BE075762
31	129.4	12.9	302	9	AI904566
32	126.2	12.8	651	9	AL595869
33	126.4	12.6	501	17	FR0007368
34	124.8	12.5	914	17	CNS01SP
35	122	12.2	841	9	AI323021
36	121	12.1	702	13	BI827961
37	117.6	11.7	613	9	AA543451
38	117.6	11.7	677	12	BE854363
39	117.6	11.7	710	12	BE854385
40	116.4	11.6	456	17	AQ586629
41	114	11.4	138	10	BE483678
42	114	11.4	347	17	BH079319
43	112.8	11.3	513	14	BQ559158
44	112.8	11.3	587	13	BJ008997
45	112.8	11.3	636	17	AG050726

## ALIGNMENTS

RESULT 1	BI016677	631 bp	linear	EST 13-JUN-2001
LOCUS	MR4-ET0138-190301-004-h03	ET0138	Homo sapiens	cdna, mRNA sequence.
DEFINITION	BI016677			
ACCESSION	BI016677.1	GI:14420748		
VERSION	EST.			
KEYWORDS	human.			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
REFERENCE	1 (bases 1 to 631)			
AUTHORS	Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.			
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags			
JOURNAL	Proc. Natl. Acad. Sci. U.S.A.	97 (7),	3491-3496	(2000)
MEDLINE	20202663			
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001			



Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR4&t2=MR4-ET0138-190301-004-h03&t3=2001-03-19&t4=1)  
Seq primer: puc 18 forward  
High quality sequence start: 19  
High quality sequence stop: 551.

## FEATURES

Location/Qualifiers  
1..631  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="ET0138"  
/dev\_stage="Adult"  
/note="Organ: lung\_tumor; Vector: puc18; Site\_1: Sma1;  
Site\_2: Sma1; A mini-library was made by cloning products  
derived from ORESTES PCR (U.S. Letters Patent application  
No. 196,716 - Ludwig Institute for Cancer Research)  
profiles into the puc 18 vector. Reverse transcription of  
tissue mRNA and cDNA amplification were performed under  
low stringency conditions."

114 a 160 c 229 g 128 t

## BASE COUNT

Query Match 36.2%; Score 362.6; DB 13; Length 631;  
Best Local Similarity 98.7%; Pred. No. 2.6e-72;  
Matches 376; Conservative 0; Mismatches 4; Indels 1; Gaps 1;

QY 136 ATGCGCTGGCTATGGGTGGGGTGGTGTGTCAGGGCCACAGCTGTGCTCATGGG 195  
|||||  
DB 205 ATGCGGTGGCTATGGGTGGGGTGGGGTGTGTCAGGGCCACAGCTGTGCTCATGGG 264  
QY 196 GCTTCTGGGGCAGAACTTGATGTGGTGGTGGGATGAGGGCTGGAGTGCCTGGC 255  
|||||  
DB 265 GCTTCTGGGGCAGAACTTGATGTGGTGGTGGGATGAGGGCTGGAGTGCCTGGC 324  
QY 256 ATGCGTTCCTCCCGTGAACCGCTGCTGTGCGCGTCTTACAGCTGGTGGTGCACCT 315  
|||||  
DB 325 AATGCTTCCTCCCGTGAACCGCTGCTGTGCGCGTCTTACAGCTGGTGGTGCACCT 384  
QY 316 CTTACAGCTGGCAGCTCTCTGTACAGCTGGGGTGGGAGAGCGCTGTGAGCGCTGCA 375  
DB 385 CTTACAGCTGGCAGCTCTCTGTACAGCTGGGGTGGGAGAGCGCTGTGAGCGCTGCA 444  
QY 376 GCTGCCCTTTAACCCCTCTGCCCTCCACAGCTGTGAGAAAGACCGCAAGAGAA 435  
|||||  
DB 445 GCTGCCCTTTAACCCCTCTGCCCTCCACAGCTGTGAGAAAGACCGCAAGAGAA 504  
QY 436 GAAGAGGTGCCAAGCCCGAGTGTCTGTAGAGCTACAGCTGACGCCGAGGTGGGGA 495  
DB 505 GAAGAGGTGCCAAGCCCGAGTGTCTGTAGAGCTACAGCTGACGCCGAGGTGGGGA 564  
QY 496 GCTATTGAGAAGGTGCGCAA 516  
|||||

DB 565 -CTCATTTGAGATGTCGCAA 584

RESULT 2  
LOCUS BM817785 560 bp mRNA linear EST 06-MAR-2002  
DEFINITION X-EST0084263 S21SNUS20s1 Homo sapiens cDNA clone S21SNUS20s1-2-F03  
5', mRNA sequence.

ACCESSION BM817785

VERSION BM817785.1 GI:19174198

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 560)

AUTHORS Kim,N.S.; Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,  
Kim,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.

TITLE  
JOURNAL  
COMMENT

21c Frontier Korean EST Project 2001  
Unpublished (2002)  
Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoen-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: yongsung@mail.kribb.re.kr  
Plate: 2 row: F column: 03  
High quality sequence stop: 560.  
Location/Qualifiers  
1..560  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="S21SNUS20s1-2-F03"  
/clone\_lib="S21SNUS20s1"  
/sex="F"  
/tissue\_type="Stomach"  
/cell\_type="Floating aggregates"  
/cell\_line="SNU-520"  
/lab\_host="Top10F"  
/note="Organ: Stomach; Vector: pTZ18RP1; Site\_1: EcoRI;  
Site\_2: NotI; The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tabacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including EcoR  
I site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized from oligo dt-selected mRNA by  
priming with dt-tailed vector. The dt-tailed vector was  
adjusted to have about 60nt. The cDNA vector was  
circularized with E. coli DNA ligase after digestion of  
EcoRI which site is also included in vector. An RNA strand  
converted to a DNA strand by Okayama-Berg method. The  
obtained cDNA vectors were used for transformation of  
competent cells E. coli Top10F' by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library. After analyzing and  
sequencing about 2,000 - 3,000 colonies in original cDNA  
library, the abundant cDNAs were selected and amplified by  
PCR reaction using vector region primer including T7  
promotor as 5' primer and N(dt)14 as 3' primer. The PCR  
products were used as template for synthesis of  
biotinylated single stranded RNA by in vitro transcription  
reaction. The synthesized RNA probes were hybridized with  
antisense single stranded cDNAs prepared from original  
library and incubated with avidin-gel. After removing  
DNA-RNA hybrids by centrifuge, the subtracted cDNA  
libraries were constructed by transfection of the  
remaining DNA into competent cells E. coli Top10F' with  
electroporation method."

FEATURES  
source

1..560  
Location/Qualifiers  
1..560  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="S21SNUS20s1-2-F03"  
/clone\_lib="S21SNUS20s1"  
/sex="F"  
/tissue\_type="Stomach"  
/cell\_type="Floating aggregates"  
/cell\_line="SNU-520"  
/lab\_host="Top10F"

118 a 187 c 160 g 95 t

BASE COUNT  
ORIGIN

Query Match 17.4%; Score 174; DB 14; Length 560;  
Best Local Similarity 100.0%; Pred. No. 2.1e-29;  
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACAACTGTCTCTGTGACATTTGACCTCTGGGACAAAGTTTCAGTGA 887

DB 88 GAACAACAGCTCAGAACAACTGTCTCTGTGACATTTGACCTCTGGGACAAAGTTTCAGTGA 147

QY 888 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTAC 947

DB 148 ACTCTCCACCAAGTGCATCATTAAGACTGTGGAGTTGCCAAGCAGCTGCCCGGCTTAC 207

QY 948 CACCTCACCATGCCCGACAGATCACCTCTCAAGCTCCCTGCCTGGACAT 1001

DB 208 CACCTCACCATGCCCGACAGATCACCTCTCAAGCTCCCTGCCTGGACAT 261

RESULT 3  
BM790595

```

LOCUS      BM790595                646 bp      mRNA      linear      EST 05-MAR-2002
DEFINITION K-EST0070459 S21SNU520 Homo sapiens cDNA clone S21SNU520-11-G11 5',
            mRNA sequence.
ACCESSION  BM790595
VERSION     BM790595.1  GI:19138827
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 646)
AUTHORS     Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,
            Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and
            Kim,Y.S.
TITLE       21C Frontier Korean EST Project 2001
JOURNAL     Unpublished (2002)
COMMENT     Contact: Kim YS
            Genome Research Center
            Korea Research Institute of Bioscience & Biotechnology
            52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea
            Tel: +82-42-860-4470
            Fax: +82-42-860-4409
            Email: yongsung@mail.kribb.re.kr
            Plate: 11 row: G column: 11
            High quality sequence stop: 646.
FEATURES    Location/Qualifiers
             1..646
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="S21SNU520-11-G11"
                /clone_lib="S21SNU520"
                /sex="F"
                /tissue_type="Stomach"
                /cell_type="Floating aggregates"
                /lab_host="Top10p"
                /note="Organ: Stomach; Vector: pTV18Rpl; Site_1: EcoRI;
                Site_2: NotI; The poly (A)+ RNA was dephosphorylated with
                bacterial alkaline phosphatase (BAP) and then decapped
                with tabacco acid pyrophosphatase (TAP). The decapped
                intact mRNA was ligated with DNA-RNA linker including EcoR
                I site by treatment of T4 RNA ligase and the first strand
                cDNA was synthesized from oligo dt-selected mRNA by
                priming with dt-tailed vector. The dt-tailed vector was
                adjusted to have about 60nt. The cDNA vector was
                circularized with E. coli DNA ligase after digestion of
                EcoRI which site is also included in vector. An RNA strand
                converted to a DNA strand by Okayama-Berg method. The
                obtained cDNA vectors were used for transformation of
                competent cells E. coli Top10p by electroporation method.
                The cDNA libraries constructed by this method are
                full-length enriched cDNA library."
BASE COUNT  140 a 213 c 184 g 109 t
ORIGIN
Query Match      17.4%; Score 174; DB 14; Length 646;
Best Local Similarity 100.0%; Pred. No. 2.2e-29;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  828  GAACAACAGCTCAGAACACGTGTCTCTCGACATTACCTCTCGGACAAAGTTCAAGTGA 887
      |||||||
Db   87  GAACAACAGCTCAGAACACGTGTCTCTCGACATTACCTCTCGGACAAAGTTCAAGTGA 146

QY  888  ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAC 947
      |||||||
Db   147  ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAC 206

QY  948  CACCTCCACCATCGCCGACCAAGATCACCTCTCTCAAGCTGCCTCGCTGGACAT 1001
      |||||||
Db   207  CACCTCCACCATCGCCGACCAAGATCACCTCTCTCAAGCTGCCTCGCTGGACAT 260

LOCUS      BE794377                982 bp      mRNA      linear      EST 20-SEP-2000
DEFINITION 601589560F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3943719 5',
            mRNA sequence.
ACCESSION  BE794377
VERSION     BE794377.1  GI:10215562
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

```

BQ706025                885 bp      mRNA      linear      EST 16-JUL-2002
LOCUS      AGENCOURT_8351580 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6282394
DEFINITION 5', mRNA sequence.
ACCESSION  BQ706025
VERSION     BQ706025.1  GI:21844924
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1 (bases 1 to 885)
AUTHORS     NIH-MGC http://mgi.nci.nih.gov/
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished (1999)
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgabbs-remail.nih.gov
            Tissue Procurement: Dr. Mark Watson
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Agencourt Bioscience Corporation
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LCM2476 row: 1 column: 11
            High quality sequence stop: 599.
FEATURES    Location/Qualifiers
             1..885
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:6282394"
                /clone_lib="NIH_MGC_113"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: spleen; Vector: pOTB7; Site_1: XhoI; Site_2:
                EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GCACGAG(G). Library constructed by Ling Hong in the
                laboratory of Gerald M. Rubin (University of California,
                Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                Superscript II RT (Life Technologies). Note: this is a
                NIH-MGC Library."
BASE COUNT  157 a 337 c 234 g 154 t 3 others
ORIGIN
Query Match      17.4%; Score 174; DB 14; Length 885;
Best Local Similarity 100.0%; Pred. No. 2.3e-29;
Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  828  GAACAACAGCTCAGAACACGTGTCTCTCGACATTACCTCTCGGACAAAGTTCAAGTGA 887
      |||||||
Db   64  GAACAACAGCTCAGAACACGTGTCTCTCGACATTACCTCTCGGACAAAGTTCAAGTGA 123

QY  888  ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAC 947
      |||||||
Db   124  ACTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAC 183

QY  948  CACCTCCACCATCGCCGACCAAGATCACCTCTCTCAAGCTGCCTCGCTGGACAT 1001
      |||||||
Db   184  CACCTCCACCATCGCCGACCAAGATCACCTCTCTCAAGCTGCCTCGCTGGACAT 237

RESULT  5
BE794377
LOCUS      601589560F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3943719 5',
DEFINITION mRNA sequence.
ACCESSION  BE794377
VERSION     BE794377.1  GI:10215562
KEYWORDS    EST.
SOURCE      human.
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

# REFERENCE

## AUTHORS

### TITLE

#### JOURNAL

##### COMMENT

1 (bases 1 to 982)  
 NIH-MGC <http://mgc.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
 Tissue Procurement: DCTD/DTF  
 cDNA Library Preparation: Ling Hong/Rubin Laboratory  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at: [image.llnl.gov](http://image.llnl.gov)  
 Plate: LCM799 row: g column: 16  
 High quality sequence stop: 860.

# FEATURES

## source

1. 982  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3943719"  
 /clone\_lib="NIH\_MGC\_7"  
 /tissue\_type="small cell carcinoma"  
 /cell\_line="MGC3"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lung; Vector: pOTB7; Site\_1: XhoI; Site\_2:  
 EcoRI; cDNA made by oligo-dT priming. Directionally  
 cloned into EcoRI/XhoI sites using the following 5'  
 adaptor: GGCACGAG(G). Size-selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."

BASE COUNT 231 a 283 c 317 g 151 t  
 ORIGIN  
 Query Match 17.4%; Score 174; DB 12; Length 982;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-29;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 828 GAACAACAGCTCAGAACACGTGTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 887  
 Db 381 GAACAACAGCTCAGAACACGTGTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 440  
 QY 888 ACTCTCCACCAAGTCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 947  
 Db 441 ACTCTCCACCAAGTCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 500  
 QY 948 CACCCTCACCATCGCGGACCATCACCTCTCTCAAGGCTGCTGCTGGACAT 1001  
 Db 501 CACCCTCACCATCGCGGACCATCACCTCTCTCAAGGCTGCTGCTGGACAT 554

# RESULT 6

## AL556392

### LOCUS

#### DEFINITION

##### ACCESSION

##### VERSION

##### KEYWORDS

##### SOURCE

##### ORGANISM

AL556392 992 bp mRNA linear EST 16-FEB-2001  
 AL556392 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CS0DK004YG05 5  
 prime, mRNA sequence.  
 AL556392  
 AL556392.1 GI:12899025  
 EST.  
 human.

# REFERENCE

## AUTHORS

### TITLE

#### JOURNAL

##### COMMENT

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 Li, W.B., Gruber, C., Jessee, J., and Polayes, D.  
 Full-length cDNA libraries and normalization  
 Unpublished (2001)  
 Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: [seqref@genoscope.cns.fr](mailto:seqref@genoscope.cns.fr), Web : [www.genoscope.cns.fr](http://www.genoscope.cns.fr).

# FEATURES

## source

1. 992  
 Location/Qualifiers

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0DK004YG05"  
 /clone\_lib="LTI\_NFL006\_PL2"  
 /tissue\_type="placenta"  
 /note="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA  
 was primed with a NotI-oligo(dT) primer. Five prime end  
 enriched, double-stranded cDNA was digested with Not I and  
 cloned into the Not I and Eco RV sites of the pCMVSPORT 6  
 vector. Library was normalized. Library was constructed by  
 Life Technologies. Contact : Feng Liang Life Technologies,  
 a division of Invitrogen 9800 Medical Center Drive  
 Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
 Email : [fliang@lifetech.com](mailto:fliang@lifetech.com) URL :  
<http://fulllength.invitrogen.com>

BASE COUNT 228 a 319 c 269 g 173 t  
 ORIGIN

Query Match 17.4%; Score 174; DB 9; Length 992;  
 Best Local Similarity 100.0%; Pred. No. 2.3e-29;  
 Matches 174; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGTGTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 887  
 Db 488 GAACAACAGCTCAGAACACGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 547  
 QY 888 ACTCTCCACCAAGTCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 947  
 Db 548 ACTCTCCACCAAGTCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGGCTTCAC 607  
 QY 948 CACCCTCACCATCGCGGACCATCACCTCTCTCAAGGCTGCTGCTGGACAT 1001  
 Db 608 CACCCTCACCATCGCGGACCATCACCTCTCTCAAGGCTGCTGCTGGACAT 661

# RESULT 7

## BI907041

### LOCUS

#### DEFINITION

##### ACCESSION

##### VERSION

##### KEYWORDS

##### SOURCE

##### ORGANISM

BI907041 700 bp mRNA linear EST 16-OCT-2001  
 603065004F1 NIH\_MGC\_118 Homo sapiens cDNA clone IMAGE:5214000 5',  
 mRNA sequence.  
 BI907041  
 BI907041.1 GI:16169824  
 EST.  
 human.

# REFERENCE

## AUTHORS

### TITLE

#### JOURNAL

##### COMMENT

Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 NIH-MGC <http://mgc.nci.nih.gov/>  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: [cgabbs@mail.nih.gov](mailto:cgabbs@mail.nih.gov)  
 Tissue Procurement: Life Technologies, Inc.  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: LLAM11537 row: h column: 01  
 High quality sequence stop: 623.

# FEATURES

## source

1. 700  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5214000"  
 /clone\_lib="NIH\_MGC\_118"  
 /tissue\_type="leukocyte"  
 /lab\_host="DH10B"  
 /note="Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV  
 (destroyed); RNA source leukocytes from anonymous pool of  
 non-activated adult donors. Library is oligo-dT primed

and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 1.2-3.3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 027. Note: this is a NIH\_MGC Library."

BASE COUNT 179 a 210 c 189 g 122 t  
ORIGIN

Query Match 17.2%; Score 172.4; DB 13; Length 700;  
Best Local Similarity 99.4%; Pred. No. 5.1e-29;  
Matches 173; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGAACACGCTCTCTCTGGACATGACCTGGGACAGTTCAGTGA 887  
|||||  
Db 360 GAACAACAGCTCAGAACACGCTCTCTCTGGACATGACCTGGGACAGTTCAGTGA 419  
|||||

QY 888 ACTCTCCACCAAGTCATTAAGACCTGTGGAGTTCGCAAGCAGCTGCCCGCTTCAC 947  
|||||  
Db 420 ACTCTCCACCAAGTCATTAAGACCTGTGGAGTTCGCAAGCAGCTGCCCGCTTCAC 479  
|||||

QY 948 CACCCTCACCATCGCGGACAGACACCTCTCAAGGCTGCTGCCTGGACAT 1001  
|||||  
Db 480 CACCCTCACCATCGCGGACAGACACCTCTCAAGGCTGCTGCCTGGACAT 533  
|||||

RESULT 8  
BF848632/c  
LOCUS  
DEFINITION IL5-EN0086-061100-237-h02 EN0086 Homo sapiens cDNA, mRNA sequence.  
ACCESSION BF848632  
VERSION BF848632.1 GI:12235782  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 173)  
AUTHORS Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R., Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F., Goldman,G.H., Carvalho,A.F., Matsukuma,A., Baia,G.S., Simpson,D.H., Brunstein,A., deoliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and Simpson,A.J.J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed sequence tags  
JOURNAL Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)  
MEDLINE 20202663

COMMENT Contact: Simpson A.J.G.  
Laboratory of Cancer Genetics  
Ludwig Institute for Cancer Research  
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil

Tel: +55-11-2704922  
Fax: +55-11-2707001

Email: asimpson@ludwig.org.br  
This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL  
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=IL5&t2=IL5-EN0086-061100-237-h02&t3=2000-11-06&t4=1)

Seq primer: puc 18 forward  
High quality sequence start: 34  
High quality sequence stop: 171.

FEATURES  
source Location/Qualifiers  
1..173  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="EN0086"  
/dev\_stage="Adult"

/note="Organ: lung\_normal; Vector: puc18; Site\_1: Smal; Site\_2: Smal; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research)

profiles into the pUC 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."

BASE COUNT 33 a 51 c 49 g 40 t  
ORIGIN

Query Match 17.0%; Score 170.4; DB 12; Length 173;  
Best Local Similarity 99.4%; Pred. No. 1.2e-28;  
Matches 171; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 738 GGCTGGTAGAGGCGAGCCCTGTGGGGCTGGAGCCAGCTGAGAAGGGTGCCATGGAG 797  
|||||

Db 173 GGCTGGTAGAGGCGAGCCCTGTGGGGCTGGAGCTAGCTGAGAAGGGTGCCATGGAG 114  
|||||

QY 798 AAGAAGCCCTCACTCTCCCTCTCCCTCCAGAACACAGCTCAGAACAACGTGTCTCTCT 857  
|||||

Db 113 AAGAAGCCCTCACTCTCCCTCTCCCTCCAGAACACAGCTCAGAACAACGTGTCTCTCT 54  
|||||

QY 858 GGACATTGACCTCTGGGACAAAGTTCAGTGAACCTCCACCAGTCATCATTT 909  
|||||

Db 53 GSACATTGACCTCTGGGACAAAGTTCAGTGAACCTCCACCAGTCATCATTT 2  
|||||

RESULT 9  
R73335  
LOCUS  
DEFINITION Y110b08.r1 Soares breast 2NBHst Homo sapiens cDNA clone IMAGE:157815 5' similar to gb:M73779 RETINOIC ACID RECEPTOR ALPHA-1 (HUMAN);, mRNA sequence.  
ACCESSION R73335  
VERSION R73335.1 GI:847367  
KEYWORDS EST.  
SOURCE human.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE 1 (bases 1 to 458)  
AUTHORS Hillier,L., Clark,N., Dubuque,T., Elliston,K., Hawkins,M., Holman ,M., Hultman,M., Kucaba,T., Le,M., Lennon,G., Marra,M., Parsons,J., Rifkin,L., Rohlfing,T., Soares,M., Tan,F., Trevaskis,E., Waterston ,R., Williamson,A.; Wohldmann,P. and Wilson,R.

TITLE The WashU-Werck EST Project  
JOURNAL Unpublished (1995)  
COMMENT Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108

Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@watson.wustl.edu

Insert Size: 1583  
High quality sequence stops: 348  
Source: IMAGE Consortium, LLNL

This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1583 Std Error: 0.00  
Seq primer: M13RPI

High quality sequence stop: 348.  
Location/Qualifiers  
1..458

/organism="Homo sapiens"  
/db\_xref="GDB:571655"  
/db\_xref="taxon:9606"  
/clone\_lib="IMAGE:157815"  
/clone\_lib="Soares breast 2NBHst"  
/sex="Female"  
/dev\_stage="adult"

/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: breast; Vector: p773D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5' TGTTACCAATCTGAAGTGGAGCGCCGCTTTTCTTTTCTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I

and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 230. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 95 a 154 c 115 g 91 t 3 others  
ORIGIN

Query Match 17.0%; Score 170; DB 14; Length 458;  
Best Local Similarity 100.0%; Pred. No. 1.7e-28;  
Matches 170; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGACAACAGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 887  
|||||  
Db 50 GAACAACAGCTCAGACAACAGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 109  
|||||

QY 888 ACTCTCCACCAAGTCATCATTAAGACTGTGGATTCGCCAAGCAGCTGCCGCGCTTCAC 947  
|||||  
Db 110 ACTCTCCACCAAGTCATCATTAAGACTGTGGATTCGCCAAGCAGCTGCCGCGCTTCAC 169  
|||||

QY 948 CACCCCTCACCATCGCGACAGATCACCCCTCCTCAAGGCTGCCTGCCTGG 997  
|||||  
Db 170 CACCCCTCACCATCGCGACAGATCACCCCTCCTCAAGGCTGCCTGCCTGG 219  
|||||

RESULT 10  
R71970  
LOCUS  
DEFINITION YJ84e07.r1 Soares breast 2NBHbst Homo sapiens cDNA clone IMAGE:155460 5' similar to gb:M73779 RETINOIC ACID RECEPTOR ALPHA-1 (HUMAN);, mRNA sequence.

ACCESSION R71970  
VERSION R71970.1 GI:846002  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1. (bases 1 to 367)  
Hillier, L., Clark, N., Dubuque, T., Elliston, K., Hawkins, M., Holman, M., Hultman, M., Kucaba, T., Le, M., Lennon, G., Marra, M., Parsons, J., Rifkin, L., Rohlfing, T., Soares, M., Tan, F., Trevaskis, E., Waterston, R., Williamson, A., Wohldmann, P. and Wilson, R.  
The WashU-Merck EST Project  
Unpublished (1995)  
Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: est@wustl.wustl.edu  
Insert Size: 1457  
High quality sequence stops: 291  
Source: IMAGE Consortium, LLNL  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
Insert Length: 1457 Std Error: 0.00  
Seq primer: M13RP1  
High quality sequence stop: 291.  
Location/Qualifiers  
1..367  
/organism="Homo sapiens"  
/db\_xref="GDB:573359"  
/db\_xref="taxon:9606"  
/clone="IMAGE:155460"  
/clone\_lib="Soares breast 2NBHbst"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: breast; Vector: pT73D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors

FEATURES  
source

1..367  
/organism="Homo sapiens"  
/db\_xref="GDB:573359"  
/db\_xref="taxon:9606"  
/clone="IMAGE:155460"  
/clone\_lib="Soares breast 2NBHbst"  
/sex="female"  
/dev\_stage="adult"  
/lab\_host="DH10B (ampicillin resistant)"  
/note="Organ: breast; Vector: pT73D (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTT 3'], double-stranded cDNA was ligated to Eco RI adaptors

(Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of a modified pT73 vector (Pharmacia). Library went through one round of normalization to a Cot = 230. Library constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 81 a 132 c 80 g 70 t 4 others  
ORIGIN

Query Match 16.9%; Score 169; DB 14; Length 367;  
Best Local Similarity 99.4%; Pred. No. 2.7e-28;  
Matches 169; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 828 GAACAACAGCTCAGACAACAGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 887  
|||||  
Db 50 GAACAACAGCTCAGACAACAGTGTCTCTCTGGACATTGACCTCTGGGACAAAGTTTCAGTGA 109  
|||||

QY 888 ACTCTCCACCAAGTCATCATTAAGACTGTGGATTCGCCAAGCAGCTGCCGCGCTTCAC 947  
|||||  
Db 110 ACTCTCCACCAAGTCATCATTAAGACTGTGGATTCGCCAAGCAGCTGCCGCGCTTCAC 169  
|||||

QY 948 CACCCCTCACCATCGCGACAGATCACCCCTCCTCAAGGCTGCCTGCCTGG 997  
|||||  
Db 170 CACCCCTCACCATCGCGACAGATCACCCCTCCTCAAGGCTGCCTGCCTGG 219  
|||||

RESULT 11  
BF721043  
LOCUS  
DEFINITION mab62b03.y1 Soares\_thymus\_2NDMT Mus musculus cDNA clone IMAGE:3974740 5' similar to SW:RRA\_MOUSE P11416 RETINOIC ACID RECEPTOR ALPHA ;, mRNA sequence.

ACCESSION BF721043  
VERSION BF721043.1 GI:12022045  
KEYWORDS EST.  
SOURCE house mouse.  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.  
1. (bases 1 to 424)  
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-r@mail.nih.gov  
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.  
MG1:1474772  
Seq primer: -40RP from Gibco  
High quality sequence stop: 361.  
Location/Qualifiers  
1..424  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3974740"  
/clone\_lib="Soares\_thymus\_2NDMT"  
/sex="male"  
/tissue\_type="Thymus"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

FEATURES  
source

1..424  
/organism="Mus musculus"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3974740"  
/clone\_lib="Soares\_thymus\_2NDMT"  
/sex="male"  
/tissue\_type="Thymus"  
/dev\_stage="4 weeks"  
/lab\_host="DH10B"  
/note="Vector: pT73D-Pac (Pharmacia) with a modified polylinker; Site\_1: Not I; Site\_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTTACCAATCTGAAGTGGAGCGCGCCCTTTTTTTTTTTT 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pT73 vector. RNA provided by Dr. Bertrand Jordan. Library went through two rounds of normalization, and was constructed by Bento Soares and M.Fatima Bonaldo."

BASE COUNT 113 a 135 c 103 g 73 t  
ORIGIN



```
Db 391 GCCGACTGCAGAGTCTTTGAAGTGGCATGTCCAAAGGAGTCTGTGAGAAACGACCGAA 450
Qy 426 ACAAGAAGAAGAGGAGTCCCAAGCCGAGTGTCTCTGAGAGCTACAGCTGACGCCGG 485
Db 451 ACAAGAAGAAGAGGAGTCCCAAGCCGAGTGTCTCTGAGAGCTACAGCTGACGCCGG 510
Qy 486 AGGTGGGGAGCTCATTTGAGAGGTGCGCAAGCGCACGAGAAACCTTCCCTGCCCTCT 545
Db 511 AGGTGGGGAGCTCATTTGAGAGGTGCGCAAGCGCACGAGAAACCTTCCCTGCCCTCT 570
Qy 546 GCCAGCTGGGCAAAATACACTACG 568
Db 571 GCCAGCTGGGCAAAATACACTACG 593

RESULT 14
LOCUS BF182871 866 bp mRNA linear EST 31-OCT-2000
DEFINITION 601809396F1 NIH_MGC_18 Homo sapiens cDNA clone IMAGE:4040303 5',
            mRNA sequence.
ACCESSION BF182871
VERSION BF182871.1 GI:11061015
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 866)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: c9apbs-re@mail.nih.gov
            Tissue Procurement: DCTD/DTF/Gazdar
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM857 row: o column: 24
            High quality sequence stop: 723.
FEATURES
            Location/Qualifiers
            1..866
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:4040303"
                /clone_lib="NIH_MGC_18"
                /tissue_type="large cell carcinoma"
                /lab_host="DH10B (phage-resistant)"
                /note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
                EcoRI; cDNA made by oligo-dT priming. Directionally cloned
                into EcoRI/XhoI sites using the following 5' adaptor:
                GGCACGAG(G). Library constructed by Ling Hong in the
                laboratory of Gerald M. Rubin (University of California,
                Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
                Superscript II RT (Life Technologies). Note: this is a
                NIH_MGC Library."
BASE COUNT 215 a 251 c 264 g 136 t
ORIGIN
Query Match 16.6%; Score 166.2; DB 12; Length 866;
Best Local Similarity 88.7%; Pred. No. 1.4e-27;
Matches 180; Conservative 0; Mismatches 23; Indels 0; Gaps 0;

Qy 366 GCAGCTGCAGTGCCTTTAAACCCCTCTGCCCTCCACAGCTGTGAGAAACGACCGAA 425
Db 257 GCCGACTGCAGAGTGTGTTGAGTGGGCATGTCCAAAGGAGTCTGTGAGAAACGACCGAA 316
Qy 426 ACAAGAAGAAGAGGAGTCCCAAGCCGAGTGTCTGAGAGCTACAGCTGACGCCGG 485
Db 317 ACAAGAAGAAGAGGAGTCCCAAGCCGAGTGTCTGAGAGCTACAGCTGACGCCGG 376
```

Search completed: March 30, 2003, 13:48:26

```
Qy 486 AGCTGGGGAGCTCATTGAGAAGTGGCAAAAGCGCACGAGAAACCTTCCCTGCCCTCT 545
Db 377 AGTGGGGAGCTCATTGAGAAGTGGCAAAAGCGCACGAGAAACCTTCCCTGCCCTCT 436
Qy 546 GCCAGCTGGGCAAAATACACTACG 568
Db 437 GCCAGCTGGGCAAAATACACTACG 459

RESULT 15
LOCUS BE547412 861 bp mRNA linear EST 09-AUG-2000
DEFINITION 601072989F1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3458974 5',
            mRNA sequence.
ACCESSION BE547412
VERSION BE547412.1 GI:9776057
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 861)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
            Email: c9apbs-re@mail.nih.gov
            Tissue Procurement: AFCC
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: Incyte Genomics, Inc.
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLCM8451 row: a column: 23
            High quality sequence stop: 571.
FEATURES
            Location/Qualifiers
            1..861
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3458974"
                /clone_lib="NIH_MGC_12"
                /tissue_type="cervical carcinoma cell line"
                /lab_host="DH10B"
                /note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
                Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                Average insert size 1.4 kb. Library prepared by Life
                Technologies."
BASE COUNT 247 a 270 c 219 g 125 t
ORIGIN
Query Match 16.4%; Score 164.6; DB 10; Length 861;
Best Local Similarity 88.2%; Pred. No. 3.1e-27;
Matches 179; Conservative 0; Mismatches 24; Indels 0; Gaps 0;

Qy 366 GCAGCTGCAGTGCCTCTTAACCCCTCTGCCCTCCACAGCTGTGAGAAACGACCGAA 425
Db 120 GCCGACTGCAGAGTGTCTTTGAAGTGGGCATGTCTCCAGGAGTCTGTGAGAAACGACCGAA 179
Qy 426 ACAAGAAGAAGAGGAGTCCCAAGCCGAGTGTCTGAGAGCTACAGCTGACGCCGG 485
Db 180 ACAAGAAGAAGAGGAGTCCCAAGCCGAGTGTCTGAGAGCTACAGCTGACGCCGG 239
Qy 486 AGCTGGGGAGCTCATTGAGAAGTGGCAAAAGCGCACGAGAAACCTTCCCTGCCCTCT 545
Db 240 AGTGGGGAGCTCATTGAGACGCGTGGCAAAAGCGCACGAGAAACCTTCCCTGCCCTCT 299
Qy 546 GCCAGCTGGGCAAAATACACTACG 568
Db 300 GCCAGCTGGGCAAAATACACTACG 322
```

Job time : 1300.57 secs

---



GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:00:38 ; Search time 2122.53 seconds  
(without alignments)  
13725.086 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_15000\_16000  
Perfect score: 1001  
Sequence: 1 gcctaaattcccggttt.....ccctccctccactggagaa 1001

Scoring table: IDENTITY\_NUC  
; Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

- 1: gb.ba.\*
- 2: gb.htg.\*
- 3: gb.in.\*
- 4: gb.om.\*
- 5: gb.ov.\*
- 6: gb.pat.\*
- 7: gb.ph.\*
- 8: gb.pl.\*
- 9: gb.pr.\*
- 10: gb.ro.\*
- 11: gb.sts.\*
- 12: gb.sy.\*
- 13: gb.un.\*
- 14: gb.vi.\*
- 15: em.ba.\*
- 16: em.fun.\*
- 17: em.hum.\*
- 18: em.in.\*
- 19: em.mu.\*
- 20: em.om.\*
- 21: em.or.\*
- 22: em.ov.\*
- 23: em.pat.\*
- 24: em.ph.\*
- 25: em.pl.\*
- 26: em.ro.\*
- 27: em.sts.\*
- 28: em.un.\*
- 29: em.vi.\*
- 30: em.htg.hum.\*
- 31: em.htg.inv.\*
- 32: em.htg.other.\*
- 33: em.htg.mus.\*
- 34: em.htg.pln.\*
- 35: em.htg.rod.\*
- 36: em.htg.mam.\*
- 37: em.htg.vrt.\*
- 38: em.sy.\*
- 39: em.htgo.hum.\*
- 40: em.htgo.mus.\*
- 41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	871.8	87.1	2907	6	AX333042	AX333042 Sequence
2	871.8	87.1	2907	9	HSRRA	X06614 Human mRNA
3	871.8	87.1	2908	6	108117	108117 Sequence 1
4	871.8	87.1	2928	6	AR061122	AR061122 Sequence
5	871.8	87.1	2940	6	AR052136	AR052136 Sequence
c	871.8	87.1	138999	9	AC090426	AC090426 Homo sapi
c	871.8	87.1	173441	2	AC080112	AC080112 Homo sapi
8	870.2	86.9	2940	6	I09348	I09348 Sequence 1
9	868.4	86.8	166368	2	AC018629	AC018629 Homo sapi
c	858.2	85.7	158766	2	AC015851	AC015851 Homo sapi
c	851.4	85.1	188574	2	AC131063	AC131063 Homo sapi
c	851.2	85.0	190309	2	AC126392	AC126392 Homo sapi
c	840	83.9	187557	2	AC069032	AC069032 Homo sapi
14	689	68.8	2658	6	AR052137	AR052137 Sequence
15	682.8	68.2	1081	9	HSRRA1A8	AF088895 Homo sapi
16	648.2	64.8	3036	6	AR061121	AR061121 Sequence
17	648.2	64.8	3036	6	I96215	I96215 Sequence 52
18	648.2	64.8	3036	9	HUMPMLRAR	M73779 Human PML-R
19	647.2	64.7	2008	9	AK098172	AK098172 Homo sapi
20	645.4	64.5	2300	9	HSM805408	AL834159 Homo sapi
21	639.4	63.9	2441	9	BC008727	BC008727 Homo sapi
22	373	37.3	1920	9	HSRAR	X06538 Human mRNA
23	373	37.3	1944	9	HSU41743	U41743 Human nucle
24	373	37.3	2073	9	HSU41742	U41742 Human nucle
25	273.4	27.3	231248	2	AL591067	AL591067 Mus muscu
26	263.2	26.3	2748	10	AY046943	AY046943 Mesocrice
27	258.6	25.8	162393	2	AC111746	AC111746 Rattus no
28	203.4	20.3	2130	10	RNI15211	U15211 Rattus norv
29	184	18.4	808	11	GI0521	GI0521 human STS C
30	112.8	11.3	1284	9	S50916	S50916 PML-RAR fus
31	102.2	10.2	2061	10	S56656	S56656 retinoic ac
32	79.8	8.0	1572	12	AF242868	AF242868 Synthetic
33	79.8	8.0	2085	12	AF242867	AF242867 Synthetic
34	79	7.9	1474	10	MSRARA	Y10094 M.spretus m
35	76	7.6	93	6	AR052141	AR052141 Sequence
c	74.6	7.5	76141	2	AC024393	AC024393 Homo sapi
37	65.6	6.6	179608	2	AC128497	AC128497 Rattus no
38	63.4	6.3	125020	9	AF429315	AF429315 Homo sapi
39	63.2	6.3	194776	2	AC129706	AC129706 Rattus no
40	62.6	6.3	180109	2	AC125916	AC125916 Rattus no
41	61.4	6.1	171777	2	AC073151	AC073151 Mus muscu
42	61	6.1	194929	2	AC119084	AC119084 Rattus no
43	60.2	6.0	192926	2	AC116254	AC116254 Rattus no
44	60	6.0	1494	5	GGRAA21	X78335 G.gallus re
45	60	6.0	1552	5	GGRAA1	X73972 G.gallus RA

ALIGNMENTS

RESULT 1	AX333042	Sequence	3551 from Patent WO0194629.	2907 bp	DNA	linear	PAT 09-JAN-2002
LOCUS	AX333042	Sequence	3551 from Patent WO0194629.				
DEFINITION	AX333042	Sequence	3551 from Patent WO0194629.				
ACCESSION	AX333042	Sequence	3551 from Patent WO0194629.				
VERSION	AX333042.1	GI:18123676					
KEYWORDS	human.						
SOURCE	human.						
ORGANISM	Homo sapiens						
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;						
	Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.						
REFERENCE	1						
AUTHORS	Young, P.E., Augustus, M., Carter, K.C., Ebner, R., Endress, G.,						
	Horrigan, S., Soppet, D.R. and Weaver, Z.						
TITLE	Cancer gene determination and therapeutic screening using signature						



Db 1479 CCACCTCCCGTACCGCCACGCCACATGGACACAGCCCTCGCCCTCCGCCCGGCTTTT 1538  
QY 257 CTCCTGCTTTCTACCGACATGTACCCCGCAGCCAGCCCTGCGCCACCTGCGCTCCCG 316  
Db 1539 CTCCTGCTTTCTACCGACATGTACCCCGCAGCCAGCCCTGCGCCACCTGCGCTCCCG 1598  
QY 317 GCAGTACTGGGACCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376  
Db 1599 GCAGTACTGGGACCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658  
QY 377 AGGCTGGGCTTCAGTGACCTGCTGCTCCACAGCCCTGCGCTGAGCTGAGAGGCGGAG 436  
Db 1659 AGGCTGGGCTTCAGTGACCTGCTGCTCCACAGCCCTGCGCTGAGCTGAGAGGCGGAG 1718  
QY 437 GCAGGAACTAGTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 496  
Db 1719 GCCAGAACTAGTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1778  
QY 497 TTCATCAAGACACCCCTCTGCGCAGCTCACACATCTTTCATCACCAAGAAACGCCAGGAC 556  
Db 1779 TTCATCAAGACACCCCTCTGCGCAGCTCACACATCTTTCATCACCAAGAAACGCCAGGAC 1838  
QY 557 TTGGCTCCCGATCTCAGAACTCAGAGCCATTTGCTCCCGAGCTGGGAACTCAACCT 616  
Db 1839 TTGGCTCCCGATCTCAGAACTCAGAGCCATTTGCTCCCGAGCTGGGAACTCAACCT 1898  
QY 617 CCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 876  
Db 1899 CCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1958  
QY 677 ATACCTGCCATPACCAACCCCAAGGATTAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 736  
Db 1959 ATACCTGCCATPACCAACCCCAAGGATTAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTG 2018  
QY 737 TTTGCTTTGATTTTAAAGAAATTTTCAATTTTAAAGCAATTTATATCAAGGAAT 796  
Db 2019 TTTGCTTTGATTTTAAAGAAATTTTCAATTTTAAAGCAATTTATATCAAGGAAT 2078  
QY 797 TGTGCTGTATTTGGGGAGCTGATCCAGAGCTGGAGGCTGGGCTGGGCTGGGCTGGGCTGGG 856  
Db 2079 TGTGCTGTATTTGGGGAGCTGATCCAGAGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGG 2138  
QY 917 CAGCTTTTCTCCTCAGTTTCTTTAAACTGTGAAGTACTAACTTTTCAAGGCTG 976  
Db 2199 CAGCTTTTCTCCTCAGTTTCTTTAAACTGTGAAGTACTAACTTTTCAAGGCTG 2258  
QY 977 CTTTCCCTCCCTCCACTGGAGAA 1001  
Db 2259 CTTTCCCTCCCTCCACTGGAGAA 2283

RESULT 3  
I08117  
LOCUS I08117 2908 bp. DNA linear PAT 02-DEC-1994  
DEFINITION Sequence 1 from Patent EP 0325849.  
ACCESSION I08117  
VERSION I08117.1 GI:589170  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 2908)  
AUTHORS Evans,R.M., Giguere,V., Ong,E.S., Segui,P.S., Umesono,K. and Thompson,C.C.  
TITLE Retinoic acid receptor composition and method for identifying ligands  
JOURNAL Patent: EP 0325849-A2 1 02-AUG-1989;  
FEATURES Location/Qualifiers  
1. 2908

BASE COUNT 562 a 1012 c 788 g 546 t  
ORIGIN  
Query Match 87.1%; Score 871.8; DB 6; Length 2908;  
Best Local Similarity 94.4%; Pred. No. 6.3e-191;  
Matches 873; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 77 CGTNN 136  
Db 1359 CCTGGACACTCTAGCGGACAGCGCGGGGGTGGGGGGGCGGACGGGGTGGCTTGCGCCC 1418  
QY 137 CCGCGCAGGAGCTGTAGCCGCCAGCCCTCAGCCCCAGCTCCCAACAGAGCAGCGCGGCCAC 196  
Db 1419 CCGCGCAGGAGCTGTAGCCGCCAGCCCTCAGCCCCAGCTCCCAACAGAGCAGCGCGGCCAC 1478  
QY 197 CCATCTCCCGTACCGCCCGCCACGCACATGGACACAGCCCTCGCCCTCGCGCCCGGCTTTT 256  
Db 1479 CCATCTCCCGTACCGCCCGCCACGCACATGGACACAGCCCTCGCCCTCGCGCCCGGCTTTT 1538  
QY 257 CTCCTGCTTTTACCGACCATGTGACCCCGCAGCCCTGCCCCCACTGCCCCCTCCCGG 316  
Db 1539 CTCCTGCTTTTACCGACCATGTGACCCCGCAGCCCTGCCCCCACTGCCCCCTCCCGG 1598  
QY 317 GCAGTACTGGGAGCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376  
Db 1599 GCAGTACTGGGAGCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658  
QY 377 AGGCTGGGCGCTCAGTGGAGCTGCTGCTGCCACAGCGCTGGGCTGACGTCAAGAGCGCGAG 436  
Db 1659 AGGCTGGGCGCTCAGTGGAGCTGCTGCTGCCACAGCGCTGGGCTGACGTCAAGAGCGCGAG 1718  
QY 437 GCCAGGAATGAGTGGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 496  
Db 1719 GCCAGGAATGAGTGGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1778  
QY 497 TTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCAACAGCAATTTATATCAAGGAAT 556  
Db 1779 TTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCAACAGCAATTTATATCAAGGAAT 1838  
QY 557 TTGGCTCCCGATCTCAGAACTCAGAGCCATTTGCTCCCGAGCTGGGAACTCAACCT 616  
Db 1839 TTGGCTCCCGATCTCAGAACTCAGAGCCATTTGCTCCCGAGCTGGGAACTCAACCT 1898  
QY 617 CCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 676  
Db 1899 CCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1958  
QY 677 ATACCTGCCATPACCAACCCCAAGGATTAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 736  
Db 1959 ATACCTGCCATPACCAACCCCAAGGATTAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2018  
QY 737 TTTTGGTTTGTATTTTAAAGAAATTTTCAATTTTAAAGCAATTTATATCAAGGAAT 796  
Db 2019 TTTTGGTTTGTATTTTAAAGAAATTTTCAATTTTAAAGCAATTTATATCAAGGAAT 2078  
QY 797 TGTGCTGTATTTGGGGAGCTGATCCAGAGCTGGAGGCTGGGCTGGGCTGGGCTGGGCTGGG 856  
Db 2079 TGTGCTGTATTTGGGGAGCTGATCCAGAGCTGGAGGCTGGAGGCTGGAGGCTGGAGGCTGG 2138  
QY 857 GTGGCTCGGAAGGGGCGCCCACTCTCTTTTCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 916  
Db 2139 GTGGCTCGGAAGGGGCGCCCACTCTCTTTTCATGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2198  
QY 917 GAGCTTTTCTCCTCAGTTTCTTTAAACTGTGAAGTACTAACTTTTCAAGGCTG 976  
Db 2199 GAGCTTTTCTCCTCAGTTTCTTTAAACTGTGAAGTACTAACTTTTCAAGGCTG 2258  
QY 977 CTTTCCCTCCCTCCACTGGAGAA 1001  
Db 2259 CTTTCCCTCCCTCCACTGGAGAA 2283

[illegible][illegible]





















```
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
*
* 30513: contig of 30513 bp in length
* 30514 30613: gap of 100 bp
* 30614 31764: contig of 1151 bp in length
* 31765 31864: gap of 100 bp
* 31865 33693: contig of 1829 bp in length
* 33694 33793: gap of 100 bp
* 33794 33405: contig of 1612 bp in length
* 35406 35505: gap of 100 bp
* 35506 37691: contig of 2186 bp in length
* 37692 37791: gap of 100 bp
* 37792 51370: contig of 13579 bp in length
* 51371 51470: gap of 100 bp
* 51471 66014: contig of 14544 bp in length
* 66015 66114: gap of 100 bp
* 66115 108183: contig of 42069 bp in length
* 108184 108283: gap of 100 bp
* 108284 130510: contig of 22227 bp in length
* 130511 130610: gap of 100 bp
* 130611 166740: contig of 36130 bp in length
* 166741 166840: gap of 100 bp
* 166841 190309: contig of 23469 bp in length.
*
FEATURES             Location/Qualifiers
     source            1..190309
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="17"
                        /map="17"
     misc_feature       /clone="RP11-1029F16"
                        /clone_lib="RPC1-11 Human Male BAC"
                        1..30513
                        /note="assembly_fragment"
                        clone_end:SP6
                        vector_side:left
     misc_feature       30614..31764
                        /note="assembly_fragment"
                        31865..33693
                        /note="assembly_fragment"
                        33794..35405
                        /note="assembly_fragment"
                        35506..37691
                        /note="assembly_fragment"
                        37792..51370
                        /note="assembly_fragment"
                        51471..66014
                        /note="assembly_fragment"
                        66115..108183
                        /note="assembly_fragment"
                        108284..130510
                        /note="assembly_fragment"
                        130611..166740
                        /note="assembly_fragment"
                        166841..190309
                        /note="assembly_fragment"
                        clone_end:T7
                        vector_side:right
BASE COUNT  48207 a 48484 c 47543 g 45018 t 1057 others
ORIGIN
Query Match      85.0%; Score 851.2; DB 2; Length 190309;
Best Local Similarity 98.7%; Pred. No. 7.9e-186;
Matches 856; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
QY 135 CCCCCCGCAGCGCTGTAGCCCGCAGCTCAGCCCCAGCTCAACAGAGAGCGCCGGCC 194
      |||||
Db 130510 CCCCCCGCAGGGNNAANTACCCAGCCTCAGCCCCAGCTCAACAGAGAGCGCCGGCC 130451
QY 195 ACCCAGTCCCGTACCCCGCAGCCACATGGACACAGCCCTCGCCCTCCGCCCGCGGCTT 254
      |||||
Db 130450 ACCCAGTCCCGTACCCCGCAGCCACATGGACACAGCCCTCGCCCTCCGCCCGCGGCTT 130991
```

```
QY 255 TTCTCTGCCTTTCTACCGACCATGTGACCCCGCACACAGCCTGCCCCACACTGCCCCCTCCC 314
      |||||
Db 130390 TTCTCTGCCTTTCTACCGACCATGTGACCCCGCACACAGCCTGCCCCACACTGCCCCCTCCC 130331
QY 315 GGCAGTACTGGGGACCTTCCCTGGGGGACGGGGAGGAGGAGCAGCAGCTCTCTTGAC 374
      |||||
Db 130330 GGCAGTACTGGGGACCTTCCCTGGGGGACGGGGAGGAGGAGCAGCAGCTCTCTTGAC 130271
QY 375 AGAGGCTTGGGCGCTCAGTGGAGTGCCTGCTCCACAGCCTGGGCTGACGTACAGGCGC 434
      |||||
Db 130270 AGAGGCTTGGGCGCTCAGTGGAGTGCCTGCTCCACAGCCTGGGCTGACGTACAGGCGC 130211
QY 435 AGGCCAGGAACACTGAGTGAGGCCCTGCTGCTGGTCTCAGATGGGTCCCTGGGGGCGCTCG 494
      |||||
Db 130210 AGGCCAGGAACACTGAGTGAGGCCCTGCTGCTGGTCTCAGATGGGTCCCTGGGGGCGCTCG 130151
QY 495 TGTTTCATCAAGACACACCCCTCTGCCAGCTCACCACATCTTTCATCACCAGCAAAAGCCAGG 554
      |||||
Db 130150 TGTTTCATCAAGACACACCCCTCTGCCAGCTCACCACATCTTTCATCACCAGCAAAAGCCAGG 130091
QY 555 ACTTGGCTCCCCATCCTCAGAACTCACAAGCCATTGCTCCCCAGCTGGGGAACCTCAAC 614
      |||||
Db 130090 ACTTGGCTCCCCATCCTCAGAACTCACAAGCCATTGCTCCCCAGCTGGGGAACCTCAAC 130031
QY 615 CTCCCCCCTGCTCGGTGGTGACAGAGGGGTGGGAGGGGGGGGGGGGGGGGGGGGGGGTTC 674
      |||||
Db 130030 CTCCCCCCTGCTCGGTGGTGACAGAGGGGTGGGAGGGGGGGGGGGGGGGGGGGGGGGTTC 129971
QY 675 ACATACCCCTGCCATACCAACCCAGGTATTAAATTCGCTGGTTCCTTTTATTATTAAT 734
      |||||
Db 129970 ACATACCCCTGCCATACCAACCCAGGTATTAAATTCGCTGGTTCCTTTTATTATTAAT 129911
QY 735 TTTTGTGTTTGAATTTTAAAGAAATTTTCAATTTTAAAGCAGATTTATCTGAAGGAA 794
      |||||
Db 129910 TTTTGTGTTTGAATTTTAAAGAAATTTTCAATTTTAAAGCAGATTTATCTGAAGGAA 129851
QY 795 TTTGCTGCTGTATTGGGGAGCTGGATCCAGAGCTGGAGGGGGGGGGGGGGGGGGGGGGAGG 854
      |||||
Db 129850 TTTGCTGCTGTATTGGGGAGCTGGATCCAGAGCTGGAGGGGGGGGGGGGGGGGGGGGGAGG 129791
QY 855 GAGTGGCTCGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 914
      |||||
Db 129790 GAGTGGCTCGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 129731
QY 915 CTCAGCCTTTTCTCTCTCAGTTCCTTTTAAAGAACTGGAAGTACTAACTTTTCAAGGCC 974
      |||||
Db 129730 CTCAGCCTTTTCTCTCTCAGTTCCTTTTAAAGAACTGGAAGTACTAACTTTTCAAGGCC 129671
QY 975 TGCCCTTCCCTCCCTCCCACTGGAGAA 1001
      |||||
Db 129670 TGCCCTTCCCTCCCTCCCACTGGAGAA 129644

RESULT 13
AC069032/c 187557 bp DNA linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 18 clone RP11-185D6 map 18, WORKING DRAFT
DEFINITION SEQUENCE, 44 unordered pieces.
ACCESSION AC069032
VERSION AC069032.4 GI:12963008
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 187557)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abrahram,H., Allen,N.,
TITLE Homo sapiens chromosome 18, clone RP11-185D6
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 187557)
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abrahram,H., Allen,N.,
TITLE Homo sapiens chromosome 18, clone RP11-185D6
JOURNAL Unpublished
AUTHORS Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,
```

Campopiano, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collimore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S.,  
 Dodge, S., Domino, M., Doyle, M., Ferrelira, P., Fitzhugh, W., Gage, D.,  
 Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
 Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L.,  
 Howland, J.C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A.,  
 Klein, J., LaRocque, K., Lamazares, R., Landers, T., Lehoczky, J.,  
 Levine, R., Liu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N.,  
 McCarthy, M., McEwan, P., McGurk, A., McKernan, K., McPheeters, R.,  
 Meldrim, J., Meneus, L., Mihova, T., Miranda, C., Mienga, V., Morrow, J.,  
 Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neill, D., Ollivar, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

Direct Submission  
 Submitted (17-MAY-2000) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 3 (bases 1 to 187557)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,  
 Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B., Brown, A.,  
 Camarata, J., Campopiano, A., Choepel, Y., Colangelo, M., Collins, S.,  
 Collimore, A., Cooke, P., DeArellano, K., Dewar, K., Diaz, J.S.,  
 Dodge, S., Faro, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J.,  
 Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
 Hagos, B., Heaford, A., Horton, L., Hulme, W., Iliev, I., Johnson, R.,  
 Jones, C., Karatas, A., LaRocque, K., Lamazares, R., Landers, T.,  
 Lehoczky, J., Levine, R., Liu, G., MacLean, C., Macdonald, P.,  
 Marquis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K.,  
 McPheeters, R., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,  
 Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C.H.,  
 O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,  
 Phunkhang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R.,  
 Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rosetti, M.,  
 Roy, A., Santos, R., Schauer, S., Schuback, R., Seaman, S., Severy, P.,  
 Sougne, C., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
 Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
 Travers, M., Travis, N., Triggillo, J., Vassiliev, H., Viel, R., Vo, A.,  
 Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G., Zainoun, J.,  
 Zembek, L., Zimmer, A. and Zody, M.

## TITLE

## JOURNAL

Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome  
 Research, 320 Charles Street, Cambridge, MA 02141, USA  
 On Feb 18, 2001 this sequence version replaced gi:11225387.  
 All repeats were identified using RepeatMasker:  
 Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

## COMMENT

----- Genome Center  
 Center: Whitehead Institute/ MIT Center for Genome Research  
 Center code: WIBR  
 Web site: <http://www-seq.wi.mit.edu>  
 Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
 ----- Project Information  
 Center project name: Li0383  
 Center clone name: 185\_D\_6

## ----- Summary Statistics

Sequencing vector: M13; M7815; 2% of reads  
 Sequencing vector: Plasmid; n/a; 98% of reads  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 166116 bases at least Q40  
 Consensus quality: 176954 bases at least Q30  
 Consensus quality: 180848 bases at least Q20  
 Insert size: 167000; agarose-IP  
 Insert size: 183257; sum-of-contigs  
 Quality coverage: 3.7 in Q20 bases; agarose-IP  
 Quality coverage: 3.4 in Q20 ba.

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 44 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 1294: contig of 1294 bp in length  
 \* 1295 1394: gap of 100 bp  
 \* 1395 2505: contig of 1111 bp in length  
 \* 2506 2605: gap of 100 bp  
 \* 2606 3992: contig of 1387 bp in length  
 \* 3993 4092: gap of 100 bp  
 \* 4093 5527: contig of 1435 bp in length  
 \* 5528 5627: gap of 100 bp  
 \* 5628 6675: contig of 1048 bp in length  
 \* 6676 6775: gap of 100 bp  
 \* 6776 8026: contig of 1251 bp in length  
 \* 8027 8126: gap of 100 bp  
 \* 8127 9162: contig of 1036 bp in length  
 \* 9163 9262: gap of 100 bp  
 \* 9263 10351: contig of 1089 bp in length  
 \* 10352 10451: gap of 100 bp  
 \* 10452 11757: contig of 1306 bp in length  
 \* 11758 11857: gap of 100 bp  
 \* 11858 12878: contig of 1021 bp in length  
 \* 12879 12978: gap of 100 bp  
 \* 12979 14159: contig of 1181 bp in length  
 \* 14160 14259: gap of 100 bp  
 \* 14260 15818: contig of 1559 bp in length  
 \* 15819 15918: gap of 100 bp  
 \* 15919 17145: contig of 1227 bp in length  
 \* 17146 17245: gap of 100 bp  
 \* 17246 18765: contig of 1520 bp in length  
 \* 18766 18865: gap of 100 bp  
 \* 18866 20138: contig of 1273 bp in length  
 \* 20139 20238: gap of 100 bp  
 \* 20239 21879: contig of 1641 bp in length  
 \* 21880 21979: gap of 100 bp  
 \* 21980 23877: contig of 1898 bp in length  
 \* 23878 23977: gap of 100 bp  
 \* 23978 25680: contig of 1703 bp in length  
 \* 25681 25780: gap of 100 bp  
 \* 25781 28099: contig of 2319 bp in length  
 \* 28100 28199: gap of 100 bp  
 \* 28200 30797: contig of 2598 bp in length  
 \* 30798 30897: gap of 100 bp  
 \* 30898 33191: contig of 2294 bp in length  
 \* 33192 33291: gap of 100 bp  
 \* 33292 36520: contig of 3229 bp in length  
 \* 36521 36620: gap of 100 bp  
 \* 36621 38842: contig of 2222 bp in length  
 \* 38843 38942: gap of 100 bp  
 \* 38943 42470: contig of 3528 bp in length  
 \* 42471 42570: gap of 100 bp  
 \* 42571 45581: contig of 3011 bp in length  
 \* 45582 45681: gap of 100 bp  
 \* 45682 48005: contig of 2324 bp in length  
 \* 48006 48105: gap of 100 bp  
 \* 48106 51907: contig of 3802 bp in length  
 \* 51908 52007: gap of 100 bp  
 \* 52008 55311: contig of 3304 bp in length  
 \* 55312 55411: gap of 100 bp  
 \* 55412 59353: contig of 3942 bp in length  
 \* 59354 59453: gap of 100 bp  
 \* 59454 63506: contig of 4053 bp in length  
 \* 63507 63606: gap of 100 bp  
 \* 63607 68052: contig of 4446 bp in length  
 \* 68053 68152: gap of 100 bp  
 \* 68153 72726: contig of 4574 bp in length  
 \* 72727 72826: gap of 100 bp  
 \* 72827 78593: contig of 5767 bp in length  
 \* 78594 78693: gap of 100 bp  
 \* 78694 84391: contig of 5698 bp in length  
 \* 84392 84491: gap of 100 bp

	* 8492	92490:	contig of 7999 bp in length	
*	92491	92590:	gap of 100 bp	
*	92591	99730:	contig of 7140 bp in length	
*	99731	99830:	gap of 100 bp	
*	99831	107602:	contig of 7772 bp in length	
*	107603	107702:	gap of 100 bp	
*	107703	118403:	contig of 8701 bp in length	
*	116404	116503:	gap of 100 bp	
*	116504	126259:	contig of 9756 bp in length	
*	126260	126359:	gap of 100 bp	
*	126360	136166:	contig of 9807 bp in length	
*	136167	136286:	gap of 100 bp	
*	136267	143578:	contig of 9312 bp in length	
*	143579	145678:	gap of 100 bp	
*	145679	157186:	contig of 11508 bp in length	
*	157187	157286:	gap of 100 bp	
*	157287	170785:	contig of 13499 bp in length	
*	170786	170885:	gap of 100 bp	
*	170886	187557:	contig of 16672 bp in length.	
FEATURES				
source				
1..187557				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/chromosome="18"				
/map="18"				
/clone="RP11-185D6"				
/clone_lib="RPC1-11 Human Male BAC"				
1..1294				
misc_feature				
/note="assembly_fragment"				
1395..2505				
misc_feature				
/note="assembly_fragment"				
2606..3992				
misc_feature				
/note="assembly_fragment"				
4093..5527				
misc_feature				
/note="assembly_fragment"				
5628..6675				
<hr/>				
Query Match         83.9% ; Score 840; DB 2; Length 187557;				
Best Local Similarity   100.0%; Pred. No. 3 le-183;				
Matches 840; Conservative 0; Mismatches 0; Indels 0; Gaps 0;				
Qy	162	CTCAGCCCGGCTTCAACAGAAGCAGCCGGCGCACCCACTCCCGTGACCGGCCACGCACA	221	
Db	11755	CTCAGCCCCAGCTCCAACAGAAGCAGCCGGCGCACCCACTCCCGTGACCGGCCACGCACA	11696	
Qy	222	CATGACACACGCTCGCCCTCCGCCGGGTTCCTGCTGCTTCTACCCACCATGGA	281	
Db	11695	CATGACACACGCTCGCCCTCCGCCGGGTTCCTGCTGCTTCTACCCACCATGGA	11636	
Qy	282	CCCGCACCAGCCCTGCCCCACCTGCCCTCCCGGCGAGTAGTGGGGACCTTCCTCGGGG	341	
Db	11635	CCCGCACCAGCCCTGCCCCACCTGCCCTCCCGGCGAGTAGTGGGGACCTTCCTCGGGG	11576	
Qy	342	GACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	401	
Db	11575	GACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG	11516	
Qy	402	TGCTCCCACAGCCCTGGGCTGACGTGACAGGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGG	461	
Db	11515	TGCTCCCACAGCCCTGGGCTGACGTGACAGGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGG	11456	
Qy	462	TCCGTGGGTCTCAGGATGGGTCTCGGGGCCCTCGGTTCATCAAGACACCCCTCTGCCAAG	521	
Db	11455	TCCGTGGGTCTCAGGATGGGTCTCGGGGCCCTCGGTTCATCAAGACACCCCTCTGCCAAG	11396	
Qy	522	CTCACCACATCTTCATCACCAGCAAAACCCAGGAGCTGGCTCCCGCCATCTCAGAACTCA	581	
Db	11395	CTCACCACATCTTCATCACCAGCAAAACCCAGGAGCTGGGTCCCGCCATCTCAGAACTCA	11336	
Qy	582	CAAGCCATTGCTCCCGAGCTGGGAACTCAACCTCCCGCTCGGTGGGTGAGACA	641	
Db	11335	CAAGCCATTGCTCCCGAGCTGGGAACTCAACCTCCCGCTCGGTGGGTGAGACA	11276	



```
QY 673 GTACATACCTGCCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTTATTTA 732
Db 1673 GTACATACCTGCCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTTATTTA 1732
QY 733 ATTTTGTGTTGATTTTATTAAGAATTTTCATTTTAAGCACATTTACTGAAGG 792
Db 1733 ATTTTGTGTTGATTTTATTAAGAATTTTCATTTTAAGCACATTTACTGAAGG 1792
QY 793 AATTGTGCTGTGATTTGGGGGAGCTGGATCCAGAGCTGGAGGGGTGGTCCGGGGGA 852
Db 1793 AATTGTGCTGTGATTTGGGGGAGCTGGATCCAGAGCTGGAGGGGTGGTCCGGGGGA 1852
QY 853 GGGAGTGGCTCGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 912
Db 1853 GGGAGTGGCTCGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1912
QY 913 TCCTCAGCTTTTCCCTCCTCAGTTTCTCTTTTAAACTGTGAAGTACTAATTTCCRAAGG 972
Db 1913 TCCTCAGCTTTTCCCTCCTCAGTTTCTCTTTTAAACTGTGAAGTACTAATTTCCRAAGG 1972
QY 973 CTTGCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1001
Db 1973 CTTGCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 2001

RESULT 15
HSRARA1A8      1081 bp      DNA      linear      PRI 02-NOV-1999
LOCUS          Homo sapiens retinoic acid receptor alpha (RARA) gene, exon 9 and
DEFINITION     complete cds.
ACCESSION      AF088895
VERSION        AF088895.2 GI:6031219
KEYWORDS       8 of 8
SEGMENT        Homo sapiens.
SOURCE          Homo sapiens.
ORGANISM       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
                Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE      1 (bases 1 to 1081)
AUTHORS        Hjal, T.A. and Murray, J.C.
JOURNAL        Genomic structure of the human retinoic acid receptor-alpha gene
MEDLINE        99269925
PUBMED         10337631
REFERENCE      2 (bases 1 to 1081)
AUTHORS        Hjal, T.A. and Murray, J.C.
JOURNAL        Direct Submission
TITLE          Submitted (01-SEP-1998) Pediatrics, University of Iowa, 140 EMRB,
                Iowa City, IA 52242, USA
COMMENT        On Oct 14, 1999 this sequence version replaced gi:4160007.
FEATURES       Location/Qualifiers.
                1..1081
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /map="17q21.1"
                /clone="BAC 205M17"
                order(AF088888.2:1..1450,AF088889.2:1..1021,
                AF088890.1:1..689,AF088891.2:1..629,AF088892.1:1..1233,
                AF088893.1:1..732,AF088894.1:1..615,1..1081)
                /gene="RARA"
                join(AF088888.2:989..1192,AF088889.2:107..646,
                AF088890.1:293..441,AF088891.2:181..322,
                AF088892.1:422..582,AF088893.1:843..1019,
                AF088894.1:326..530,AF088894.1:156..314,261..1081)
                /gene="RARA"
                /product="retinoic acid receptor alpha"
                join(AF088889.2:469..646,AF088890.1:293..441,
                AF088891.2:181..322,AF088892.1:422..582,
                AF088893.1:843..1019,AF088894.1:326..530,
                AF088894.1:156..314,261..478)
                /gene="RARA"

QY 673 GTACATACCTGCCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTTATTTA 732
Db 1673 GTACATACCTGCCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTTATTTA 1732
QY 733 ATTTTGTGTTGATTTTATTAAGAATTTTCATTTTAAGCACATTTACTGAAGG 792
Db 1733 ATTTTGTGTTGATTTTATTAAGAATTTTCATTTTAAGCACATTTACTGAAGG 1792
QY 793 AATTGTGCTGTGATTTGGGGGAGCTGGATCCAGAGCTGGAGGGGTGGTCCGGGGGA 852
Db 1793 AATTGTGCTGTGATTTGGGGGAGCTGGATCCAGAGCTGGAGGGGTGGTCCGGGGGA 1852
QY 853 GGGAGTGGCTCGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 912
Db 1853 GGGAGTGGCTCGGAGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1912
QY 913 TCCTCAGCTTTTCCCTCCTCAGTTTCTCTTTTAAACTGTGAAGTACTAATTTCCRAAGG 972
Db 1913 TCCTCAGCTTTTCCCTCCTCAGTTTCTCTTTTAAACTGTGAAGTACTAATTTCCRAAGG 1972
QY 973 CTTGCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 1001
Db 1973 CTTGCTTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCC 2001

BASE COUNT 196 a 346 c 317 g 221 t 1 others
ORIGIN
Query Match 68.2%; Score 682.8; DB 9; Length 1081;
Best Local Similarity 92.9%; Pred. No. 2.7e-147;
Matches 684; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 77 CGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 136
Db 346 CCTGGACACTCTGAGCGGACAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 405
QY 137 CCCGCCAGGAGCTGTAGCCCCAGCTCAGCCCCAGCTCCACAGAGAGAGCCGCCGAC 196
Db 406 CCCGCCAGGAGCTGTAGCCCCAGCTCAGCCCCAGCTCCACAGAGAGAGCCGCCGAC 465
QY 197 CCACCTCCCGTGACCGCCAGCCACATGGACACAGAGCCCTCGCCCTCGCCCGGCTTT 256
Db 466 CCACCTCCCGTGACCGCCAGCCACATGGACACAGAGCCCTCGCCCTCGCCCGGCTTT 525
QY 257 CTCTGCTCTTTTACCGGACATGTGACCCCGACAGCCAGCTTGCCTCCAGCTGCTCCG 316
Db 526 CTCTGCTCTTTTACCGGACATGTGACCCCGACAGCCAGCTTGCCTCCAGCTGCTCCG 585
QY 317 GCAGTACTGGGACCTTCCCTGGGGGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 376
Db 586 GCAGTACTGGGACCTTCCCTGGGGGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 645
QY 377 AGGCTTGCGCTCAGTGGAGTCTGCTCCACAGAGCTTGGCTGACGAGAGGAGGAGG 436
Db 646 AGGCTTGCGCTCAGTGGAGTCTGCTCCACAGAGCTTGGCTGACGAGAGGAGGAGG 705
QY 437 GCCAGGAAGTGAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 496
Db 706 GCCAGGAAGTGAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 765
QY 497 TTCATCAAGACAGCCCTCTGCCAGCTCACCACATCTTTCATCAGCAGCAAGAGGAG 556
Db 766 TTCATCAAGACAGCCCTCTGCCAGCTCACCACATCTTTCATCAGCAGCAAGAGGAG 825
QY 557 TTGGCTCCCCCATCTCTCAGAACTCACAAGCAATTCCTCCAGCTGGGGAGACCTCAAC 616
Db 826 TTGGCTCCCCCATCTCTCAGAACTCACAAGCAATTCCTCCAGCTGGGGAGACCTCAAC 885
QY 617 CCCCGCTGCTCGTGTGGTGCAGAGGGGGTGGGACAGGGGGGGGGGGGGGGGGGGGG 676
Db 886 CCCCGCTGCTCGTGTGGTGCAGAGGGGGTGGGACAGGGGGGGGGGGGGGGGGGGGG 945
QY 677 ATACCTGCGCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTTATTTAATTT 736
Db 946 ATACCTGCGCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTTATTTAATTT 1005
QY 737 TTTTGTGTTGATTTTTTTTAAAGAAATTTTCATTTTAAGCACATTTTACTGAAGAA 796
Db 1006 TTTTGTGTTGATTTTTTTTAAAGAAATTTTCATTTTAAGCACATTTTACTGAAGAA 1065
QY 797 TGTGCTGTGTATTGGG 812
                |||||||
```

```
/codon_start=1
/product="retinoic acid receptor alpha"
/protein_id="AAD05222.1"
/db_xref="GI:4160009"
/translation="MASNSSCPPTGGGHLNGYVPPVYAFVFFPMPGLGSLPPGALTTL
QHQLPVSGYSTPSPATITETQSSSEIEIIPSPSPPPRIYKPCVQDQKSGSYHYGV
SACGCGKFFRRSIQKNMYTCHRDKNCIINKVTRNQYQRLQKCFVGNKSVYRN
DRNKKKEVPKPEGSYTLTPVEGELIEKVKARHOETFPALCOLGKYVTNNSSQRY
SLDIDLWDKFSLSLTKCIKTVDFAKQLPGTTLTADQITLLKAACTDILILICTR
YTPEDMTFTSDGLTLNRTOMHAGFGPLTDLVFAFANQLPLLEMDDETGLLSAICL
ICGDRDLEQPDPRDVMLOEPLLEALKYIVRRKRSRPHMFPMMLMKITIDLSISAKGA
ERVITLMEIPGSMPLLIQEMLENSEGLDITLSDPGPGGGGDRGGLAPPPGSCPSLSP
SSNRSSPATHSP"
261..1081
/gene="RARA"
/number=9
exon
BASE COUNT 196 a 346 c 317 g 221 t 1 others
ORIGIN
Query Match 68.2%; Score 682.8; DB 9; Length 1081;
Best Local Similarity 92.9%; Pred. No. 2.7e-147;
Matches 684; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 77 CGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 136
Db 346 CCTGGACACTCTGAGCGGACAGCGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 405
QY 137 CCCGCCAGGAGCTGTAGCCCCAGCTCAGCCCCAGCTCCACAGAGAGAGCCGCCGAC 196
Db 406 CCCGCCAGGAGCTGTAGCCCCAGCTCAGCCCCAGCTCCACAGAGAGAGCCGCCGAC 465
QY 197 CCACCTCCCGTGACCGCCAGCCACATGGACACAGAGCCCTCGCCCTCGCCCGGCTTT 256
Db 466 CCACCTCCCGTGACCGCCAGCCACATGGACACAGAGCCCTCGCCCTCGCCCGGCTTT 525
QY 257 CTCTGCTCTTTTACCGGACATGTGACCCCGACAGCCAGCTTGCCTCCAGCTGCTCCG 316
Db 526 CTCTGCTCTTTTACCGGACATGTGACCCCGACAGCCAGCTTGCCTCCAGCTGCTCCG 585
QY 317 GCAGTACTGGGACCTTCCCTGGGGGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 376
Db 586 GCAGTACTGGGACCTTCCCTGGGGGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 645
QY 377 AGGCTTGCGCTCAGTGGAGTCTGCTCCACAGAGCTTGGCTGACGAGAGGAGGAGG 436
Db 646 AGGCTTGCGCTCAGTGGAGTCTGCTCCACAGAGCTTGGCTGACGAGAGGAGGAGG 705
QY 437 GCCAGGAAGTGAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 496
Db 706 GCCAGGAAGTGAAGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 765
QY 497 TTCATCAAGACAGCCCTCTGCCAGCTCACCACATCTTTCATCAGCAGCAAGAGGAG 556
Db 766 TTCATCAAGACAGCCCTCTGCCAGCTCACCACATCTTTCATCAGCAGCAAGAGGAG 825
QY 557 TTGGCTCCCCCATCTCTCAGAACTCACAAGCAATTCCTCCAGCTGGGGAGACCTCAAC 616
Db 826 TTGGCTCCCCCATCTCTCAGAACTCACAAGCAATTCCTCCAGCTGGGGAGACCTCAAC 885
QY 617 CCCCGCTGCTCGTGTGGTGCAGAGGGGGTGGGACAGGGGGGGGGGGGGGGGGGGGG 676
Db 886 CCCCGCTGCTCGTGTGGTGCAGAGGGGGTGGGACAGGGGGGGGGGGGGGGGGGGGG 945
QY 677 ATACCTGCGCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTTATTTAATTT 736
Db 946 ATACCTGCGCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTTATTTAATTT 1005
QY 737 TTTTGTGTTGATTTTTTTTAAAGAAATTTTCATTTTAAGCACATTTTACTGAAGAA 796
Db 1006 TTTTGTGTTGATTTTTTTTAAAGAAATTTTCATTTTAAGCACATTTTACTGAAGAA 1065
QY 797 TGTGCTGTGTATTGGG 812
                |||||||
```



Db 1066 TGTCTGTGTATTGGG 1081

Search completed: March 30, 2003, 13:12:03  
Job time : 2833.53 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 07:57:38 ; Search time 196.299 Seconds  
(without alignments)  
11483.757 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_15000\_16000  
Perfect score: 1001  
Sequence: 1 gcctaaattcccggtt.....ccctccctccactggagaa 1001

Scoring table: IDENTITY\_NUC  
Gapex 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :			
N_Geneseq_101002.*			
1:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*		
2:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*		
3:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*		
4:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*		
5:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*		
6:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*		
7:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*		
8:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*		
9:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*		
10:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*		
11:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*		
12:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*		
13:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*		
14:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*		
15:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*		
16:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*		
17:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*		
18:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*		
19:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*		
20:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*		
21:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*		
22:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*		
23:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*		
24:	/SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*		

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	951	95.0	20512	24	Genomic DNA encodi
2	871.8	87.1	2907	24	Human cDNA differe
3	871.8	87.1	2907	24	Lung cancer relate
4	871.8	87.1	2928	20	Human RAR-alpha CD
5	871.8	87.1	2940	10	DNA of clone phRAR
6	871.8	87.1	2940	16	RAR-alpha clone la
7	870.2	86.9	2928	13	AAQ29338
8	842.8	84.2	2930	23	DNA encoding novel
9	689	68.8	2658	16	RAR-alpha-403 domi

10	648.8	64.8	2086	24	AAAL38338	CDNA encoding huma
11	648.2	64.8	3036	13	AAQ29334	myl/RAR-alpha fusi
12	648.2	64.8	3036	19	AAV20474	Human PML/RAR-alpha
13	648.2	64.8	3036	20	AAV64990	Human myl/RAR-alpha
14	648.2	64.8	3511	21	AAAL38656	Human PML/RAR-alpha
15	645	64.4	3511	17	AAAT33259	PML-RAR-alpha RNA
16	644.6	64.4	3511	17	AAAT33246	PML-RAR-alpha RNA
17	578	57.7	595	24	AAAS62827	CDNA sequence #614
18	565.4	56.5	3052	23	AAAS83047	DNA encoding novel
19	373	37.3	1944	23	AAAS83045	DNA encoding novel
20	373	37.3	1944	23	AAAS85136	DNA encoding novel
21	373	37.3	2073	23	AAAS85135	DNA encoding novel
22	373	37.3	2457	23	AAAS83049	DNA encoding novel
23	274.4	27.4	1481	17	AAAT33247	RAR-alpha RNA sequ
24	274.4	27.4	1481	17	AAAT33260	RAR-alpha RNA sequ
25	79.8	8.0	1692	23	AAAS83808	DNA encoding novel
26	59	5.9	296	24	ABN16897	Human ORFX polynuc
27	54.8	5.5	1000	21	AAAO2484	Human colon cancer
28	52.4	5.2	320	21	AAAS38183	Primer used in the
29	52.4	5.2	840	24	ABQ35494	Oligonucleotide fo
30	52.4	5.2	840	24	ABQ35495	Oligonucleotide fo
31	52	5.2	2188	20	AAZ77506	Human ovarian tumo
32	51.6	5.1	1337	20	AAZ17263	Human gene express
33	51.2	5.1	4475	14	AAQ45995	IDUA - exons III t
34	51.2	5.1	10732	21	AAAL10594	Gene encoding a su
35	51	5.1	600	24	ABQ52496	Oligonucleotide fo
36	51	5.1	600	24	ABQ52497	Oligonucleotide fo
37	50.6	5.1	401	22	AAI92118	Human polynucleoti
38	50.6	5.1	690	22	AAH33877	Human colon cancer
39	50.2	5.0	1327	24	ABQ68452	Listeria monocytog
40	49.6	4.9	320	21	AAAS38185	Primer used in the
41	49.4	4.9	320	21	AAAS38186	Primer used in the
42	49.4	4.9	387	23	ABV57769	Human prostate exp
43	49.4	4.9	434	22	AAI84343	Human polynucleoti
44	49	4.9	318	21	AAAS38184	Primer used in the
45	49	4.9	1232	22	AAAD11124	Human small cell l

ALIGNMENTS

RESULT 1	
AAAL38339	AAAL38339 standard; DNA; 20512 BP.
ID	AAAL38339 standard; DNA; 20512 BP.
XX	
AC	AAAL38339;
XX	
DT	15-AUG-2002 (first entry)
XX	
DE	Genomic DNA encoding the human nuclear hormone receptor protein.
DE	Human nuclear hormone receptor; therapeutic agent; gene therapy;
KW	Immune response; chromosome 17; single nucleotide polymorphism;
KW	gene; ds.
XX	
OS	Homo sapiens.
XX	
FH	Key
FT	CDS
FT	Location/Qualifiers
FT	1684..15208
FT	/*tag= a
FT	/product= "Human nuclear hormone receptor protein"
FT	1684..1846
FT	/*tag= b
FT	/number= 1
FT	1847..6808
FT	/*tag= c
FT	/number= 1
FT	6809..6957
FT	/*tag= d
FT	/number= 2
FT	6958..8277
FT	/*tag= e
FT	/number= 2

FT exon 8278..8419  
FT /tag= f  
FT /number= 3  
FT intron 8420..10406  
FT /tag= g  
FT exon 10407..10567  
FT /tag= h  
FT /number= 4  
FT intron 10568..10827  
FT /tag= i  
FT exon 10828..11004  
FT /tag= j  
FT /number= 5  
FT intron 11005..12798  
FT /tag= k  
FT exon 12799..13003  
FT /tag= l  
FT /number= 6  
FT intron 13004..13759  
FT /tag= m  
FT exon 13760..13918  
FT /tag= n  
FT /number= 7  
FT intron 13919..14505  
FT /tag= o  
FT exon 14506..14658  
FT /tag= p  
FT /number= 8  
FT intron 14659..15143  
FT /tag= q  
FT exon 15144..15208  
FT /tag= r  
FT /number= 9  
FT allele replace(4084,C)  
FT /tag= s  
FT allele replace(6482,A)  
FT /tag= t  
FT allele replace(8066,G)  
FT /tag= u  
FT allele replace(8699,C)  
FT /tag= v  
FT allele replace(12897,T)  
FT /tag= w  
FT allele replace(14442,C)  
FT /tag= x  
FT XX  
PN WO200231146-A2.  
XX  
PD 18-APR-2002.  
XX  
PF 05-OCT-2001; 2001WO-US31095.  
XX  
PR 11-OCT-2000; 2000US-239117P.  
PR 19-OCT-2000; 2000US-0691220.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Wei M, Ye J, Yan C, Ketchum KA, Di Francesco V, Beasley EM;  
XX WPI; 2002-426282/45.  
DR P-PSDB; AAO21489.  
XX  
PT New human nuclear hormone receptor proteins and nucleic acids, useful  
PT as models or targets for developing human therapeutic targets, and in  
PT identifying therapeutic proteins and modulators of nuclear hormone  
PT receptor expression  
XX

PS  
XX Claim 1; Fig 3; 73pp; English.  
CC The invention relates to an isolated peptide of a novel human nuclear  
CC hormone receptor with a fully defined sequence of 457 amino acids given  
CC in the specification. The novel human nuclear hormone receptor peptides  
CC and nucleic acids encoding them can be used as models for the development  
CC of human therapeutic targets, aid in the identification of therapeutic  
CC proteins, and serve as targets for the development of human therapeutic  
CC agents that modulate nuclear hormone receptor activity in cells and  
CC tissues that express the nuclear hormone receptor. The nucleic acids may  
CC be used as a query sequence to perform searches against sequence  
CC databases to identify family members or related sequences, as probes or  
CC primers, to construct recombinant vectors, to identify compounds that  
CC modulate nuclear hormone receptor nucleic acid expression, in gene  
CC therapy, and as antisense constructs to control nuclear hormone receptor  
CC gene expression in cells, tissues or organisms. The polypeptides can be  
CC used to raise antibodies or to elicit an immune response, as a reagent in  
CC assays designed to determine protein levels in biological fluids, as  
CC markers for tissues in which a corresponding protein is expressed, to  
CC identify a binding partner/ligand to develop a system for the  
CC identification of inhibitors of the binding reaction, in drug screening  
CC assays, and to identify compounds that modulate protein activity. This  
CC polynucleotide sequence represents the genomic DNA encoding the human  
CC nuclear hormone receptor protein of the invention.  
XX  
SQ Sequence 20512 BP; 3591 A; 5987 C; 6281 G; 4541 T; 112 other;  
Query Match 95.0%; Score 951; DB 24; Length 20512;  
Best Local Similarity 100.0%; Pred. No. 2.8e-198;  
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 GCCTAAATAATCCCGTGTGCCACCACTCTTGAAACCCCAAGCTTGGTGTATATCC 60  
Db 15000 GCCTAAATAATCCCGTGTGCCACCACTCTTGAAACCCCAAGCTTGGTGTATATCC 15059  
QY 61 CGAATTGGGGGGCCCCCGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 120  
Db 15060 CGAATTGGGGGGCCCCCGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 15119  
QY 121 NNNNNNNNNNGGGCCCCCGCCAGGAGCTGTAGCCCCAGCCTCAGCCCCAGCTCCAACA 180  
Db 15120 NNNNNNNNNNGGGCCCCCGCCAGGAGCTGTAGCCCCAGCCTCAGCCCCAGCTCCAACA 15179  
QY 181 GAAGCAGCCGGCCAGCCACTCCCGGTGACGCCCGCCACATGGACACAGCCCTCGCC 240  
Db 15180 GAAGCAGCCGGCCAGCCACTCCCGGTGACGCCCGCCACATGGACACAGCCCTCGCC 15239  
QY 241 CTCGGCCCCGGCTTTTCTCTGCCTTTCTACCGACCATGTGACCCCGCCAGCCCTCGCC 300  
Db 15240 CTCGGCCCCGGCTTTTCTCTGCCTTTCTACCGACCATGTGACCCCGCCAGCCCTCGCC 15299  
QY 301 CCACCTGCCCTCCGGGCGAGTACTGGGGAGCTTCCCTGGGGAGCGGGGAGGAGGCA 360  
Db 15300 CCACCTGCCCTCCGGGCGAGTACTGGGGAGCTTCCCTGGGGAGCGGGGAGGAGGCA 15359  
QY 361 GCGACTCCTTGGACAGAGGCTTGGGCCCTCAGTGTGCTGCTGCCACAGCTGGGCT 420  
Db 15360 GCGACTCCTTGGACAGAGGCTTGGGCCCTCAGTGTGCTGCTGCCACAGCTGGGCT 15419  
QY 421 GAGCTCAGAGGCGGAGGCCAGGAAGTGTGAGGCCCTTGGTCTCTGAGGATGGG 480  
Db 15420 GAGCTCAGAGGCGGAGGCCAGGAAGTGTGAGGCCCTTGGTCTCTGAGGATGGG 15479  
QY 481 TCCTGGGGGCTCGTGTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCAC 540  
Db 15480 TCCTGGGGGCTCGTGTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCAC 15539  
QY 541 CAGCAACGCCAGGACTTGGCTCCGCCATCCTCAGAACTCACAAGCCATTCCTCCCCAGC 600  
Db 15540 CAGCAACGCCAGGACTTGGCTCCGCCATCCTCAGAACTCACAAGCCATTCCTCCCCAGC 15599  
QY 601 TGGGGAACCTCAACCTCCCCCTGCCCTGGTGTGACAGAGGGGTGGGACAGGGCGG 660  
Db 15600 TGGGGAACCTCAACCTCCCCCTGCCCTGGTGTGACAGAGGGGTGGGACAGGGCGG 15659

Db	15600	TGGGGAACCTCAACCTCCCCCTCGCTGGGTGTGTGACAGAGGGGGTGGGACAGGGGCGG	15655
QY	661	GGGTTCCCCCTGTACATACCCCTGCCATACCAACCCAGGATTAATTTCTCGCTGGTTTT	720
Db	15660	GGGTTCCCCCTGTACATACCCCTGCCATACCAACCCAGGATTAATTTCTCGCTGGTTTT	15719
QY	721	GTTTTTAATTTAAATTTTTTGTGTTTGTGATTTTAAATAAGAAATTTTCAATTTAAGCACAT	780
Db	15720	GTTTTTAATTTAAATTTTTTGTGTTTGTGATTTTAAATAAGAAATTTTCAATTTAAGCACAT	15779
QY	781	TTATACGAAGGAATTTCTGCTGTATTTGGGGGAGCTGGATCCAGCTGGAGGGGTT	840
Db	15780	TTATACGAAGGAATTTGCTGTGTTATTTGGGGGAGCTGGATCCAGCTGGAGGGGTT	15839
QY	841	GGGTCCGGGGAGGAGCTGGCTCGAAGGGGCCCCACTCTCCTTTTCATGTCCCTGTGCC	900
Db	15840	GGGTCCGGGGAGGAGTGCTCGAAGGGGCCCCACTCTCCTTTTCATGTCCCTGTGCC	15899
QY	901	CCCCAGTTTCCTCCTCAGGCTTTTTCCTCCTCAGTTTTCCTTTTAAACTGTGAAGTACT	960
Db	15900	CCCCAGTTTCCTCCTCAGGCTTTTTCCTCCTCAGTTTTCCTTTTAAACTGTGAAGTACT	15959
QY	961	AACCTTCCAAAGGCTGCCTTCCCTCCCTCCCTCCACTGGAGAA	1001
Db	15960	AACCTTCCAAAGGCTGCCTTCCCTCCCTCCCTCCACTGGAGAA	16000
RESULT 2			
ABK84517			
ID	ABK84517 standard; cDNA; 2907 BP.		
XX	ABK84517;		
DT	14-AUG-2002 (first entry)		
XX	Human cDNA differentially expressed in granulocytic cells #1088.		
DE	Human; ss; granulocytic cell; DNA chip; bacterial infection;		
KW	viral infection; parasitic infection; protozoal infection;		
KW	fungal infection; sterile inflammatory disease; psoriasis;		
KW	rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;		
KW	cardiac reperfusion injury; renal reperfusion injury; ARDS;		
KW	adult respiratory distress syndrome; inflammatory bowel disease;		
KW	Crohn's disease; ulcerative colitis; periodontal disease;		
KW	granulocyte activation; chronic inflammation; allergy.		
OS	Homo sapiens.		
XX	WO200228999-A2.		
PN	11-APR-2002.		
PD	03-OCT-2001; 2001WO-US30821.		
XX	03-OCT-2000; 2000US-237189P.		
PF	(GENE-) GENE LOGIC INC.		
XX	Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;		
PI	WPI; 2002-435328/46.		
XX			

Detecting granulocyte activation by detecting differential expression of genes associated with granulocyte activation, which serves as diagnostic markers that is useful for monitoring disease states and drug toxicity -  
 Claim 1; SEQ ID No 1088; 114pp; English.  
 The invention relates to detecting (M1) granulocyte (GC) activation (GCA), by detecting the level of expression of gene(s) (Gs) identified by a DNA chip analysis as given in the specification, and comparing the expression level to an expression level in an unactivated granulocyte.

GC, where differential expression of Gs is indicative of GCA. Also included are modulating (M2) GA by contacting GC with an agent that alters the expression of at least one gene in Gs; (2) screening (M3) for an agent capable of modulating GCA or an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease using the gene expression profile; (3) detecting (M4) an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by detecting the level of expression in a sample of the tissue of gene(s) from Gs, where the level of expression of the gene is indicative of inflammation; (4) treating (M5) an inflammation (especially chronic) or in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease, by contacting a tissue having inflammation with an agent that modulates the expression of gene(s) from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for modulating GA; M3 is useful for screening an agent capable of modulating GCA preferably in an inflammation in a tissue; M4 is useful for detecting an inflammation (especially chronic) in a tissue, an allergic response in a subject, exposure of a subject to a pathogen or sterile inflammatory disease (e.g. psoriasis, rheumatoid arthritis, glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal reperfusion injury, ARDS, adult respiratory distress syndrome, inflammatory bowel disease, Crohn's disease, ulcerative colitis, periodontal disease; also bacterial infection, viral infection, parasitic infection, protozoal infection, fungal infection and M5 is useful for treating one of the above conditions. The present sequence represents a gene differentially expressed in granulocytes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at [ftp.wipo.int/pub/published\\_pct\\_sequences](http://wipo.int/pub/published_pct_sequences).

Sequence 2907 BP; 561 A; 1011 C; 789 G; 546 T; 0 other;

Query Match 87.1%; Score 871.8; DB 24; Length 2907;  
Best Local Similarity 94.4%; Pred. No. 3.2e-181;  
Matches 873; Conservative 0; Mismatches 52; Indels 0;

QY	77	CGTNNNGGGCCC	136
Dd	1359	CCTGGACACTCTGAGCGGACAGCGGGGGGTGGGGGGGAGCGGGTGCGCTGGCCCC	1418
QY	137	CCGCCAGGACGCTGTAGCCCCAGCCTCAGCCCCAGCTCCACACAAGCAGCCCGGCCAC	196
Dd	1419	CCGCCAGGACGCTGTAGCCCCAGCCTCAGCCCCAGCTCCAACAAGAAGCAGCCCGGCCAC	1478
QY	197	CCACTCCCGGTGACCSCCCACGCACATGGACACAGCCCTCGCCCTCCGCCCCCGGCTTTT	256
Dd	1479	CCACTCCCGGTGACCSCCCACGCACATGGACACAGCCCTCGCCCTCCGCCCCCGGCTTTT	1538
QY	257	CTCTGCTTTTCACGACCATGTGACCCCGCACAGCCCTGCCGCCACTGCCCTCCCGG	316
Dd	1539	CTCTGCTTTTCACGACCATGTGACCCCGCACAGCCCTGCCGCCACTGCCCTCCCGG	1598
QY	317	GCAGTACTGGGACCTTTCCTCGGGGAGGGGAGGAGGACGACGACTCTCTTGGACAG	376
Dd	1599	GCAGTACTGGGACCTTTCCTCGGGGAGGGGAGGAGGACGACGACTCTCTTGGACAG	1658
QY	377	AGSCCTGGGCCCTCAGTGGACTGCTGTCTCCACAGCCTGGGCTGACGTCAGAGGCCGAG	436
Dd	1659	AGSCCTGGGCCCTCAGTGGACTGCTGTCTCCACAGCCTGGCTGACGTCAGAGGCCGAG	1718
QY	437	GCCAGAACTGAGTAGGGCCCTGGTCTCTGGGTCTCAGATGGGTCTCGGGGGCTCGTG	496
Dd	1719	GCCAGAACTGAGTAGGGCCCTGGTCTCTGGGTCTCAGATGGGTCTCGGGGGCTCGTG	1778
QY	497	TTCATCAACACACCCCTCTGCCACGCTCACCACATCTTCATCACAGCAAACGCCAGAC	556
Dd	1779	TTCATCAACACACCCCTCTGCCACGCTCACCACATCTTCATCACAGCAAACGCCAGAC	1838
QY	557	TTGGCTCCCCCATCTCTCAGAACTCAAGACCATTTGCTCCCGAGCTGGGAACCTCAAGCT	616



```
QY 257 CTCGCTCTTCTACCGACATGTGACCGCGACCGAGCCCTGCCCCACCTGCCCTCCCG 316
Db 1539 CTCGCTCTTCTACCGACATGTGACCGCGACCGAGCCCTGCCCCACCTGCCCTCCCG 1598
QY 317 GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376
Db 1599 GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658
QY 377 AGGCTGGGCGCTCAGTGGAGTGCCTGTCCACAGCCCTGGGCTGACGTCAGAGGCCGAG 436
Db 1659 AGGCTGGGCGCTCAGTGGAGTGCCTGTCCACAGCCCTGGGCTGACGTCAGAGGCCGAG 1718
QY 437 GCCAGGAAGTGAAGTGGGCGGCTGTCTGGGTCTCAGAGTGGGCTGGGCGGCTCGTG 496
Db 1719 GCCAGGAAGTGAAGTGGGCGGCTGTCTGGGTCTCAGAGTGGGCTGGGCGGCTCGTG 1778
QY 497 TTCATCAAGACACCCCTGTGCCAGCTCACACATCTTCATCAGCAGCAAGCCAGGAC 556
Db 1779 TTCATCAAGACACCCCTGTGCCAGCTCACACATCTTCATCAGCAGCAAGCCAGGAC 1838
QY 557 TTGGCTCCCCATCCTCAGAACTCAAGCCATTGCTCCCCAGCTGGGGAACCTCAACCT 616
Db 1839 TTGGCTCCCCATCCTCAGAACTCAAGCCATTGCTCCCCAGCTGGGGAACCTCAACCT 1898
QY 617 CCCCCTGCCCTGCTGTGTGACAGAGGGGTGGGACAGGGGCGGGGTTCCCTCTGTAC 676
Db 1899 CCCCCTGCCCTGCTGTGTGACAGAGGGGTGGGACAGGGGCGGGGTTCCCTCTGTAC 1958
QY 677 ATACCTGCCATACCAACCCAGGATTAATTCCTGCTGGTCTGTTTATTTATTTATTT 736
Db 1959 ATACCTGCCATACCAACCCAGGATTAATTCCTGCTGGTCTGTTTATTTATTTATTT 2018
QY 737 TTTTGTGTTGATTTTAAAGAAATTTTCAATTTAAGCACATTTATATGAAGAAAT 796
Db 2019 TTTTGTGTTGATTTTAAAGAAATTTTCAATTTAAGCACATTTATATGAAGAAAT 2078
QY 797 TGTGCTGTGATTTGGGGGAGCTGATCAGAGCTGGAGGGGTGGGTCGGGGAGGGA 856
Db 2079 TGTGCTGTGATTTGGGGGAGCTGATCAGAGCTGGAGGGGTGGGTCGGGGAGGGA 2138
QY 857 GTGCTCGAAGGGGCGCCACTCCTTCATGTCCTGTGCGCCCGCCAGTCTCTCCT 916
Db 2139 GTGCTCGAAGGGGCGCCACTCCTTCATGTCCTGTGCGCCCGCCAGTCTCTCCT 2198
QY 917 CAGCCTTTTCTCCTCAGTCTTCTTTAAACTGTGAAGTACTAATTTTCAAGGCCG 976
Db 2199 CAGCCTTTTCTCCTCAGTCTTCTTTAAACTGTGAAGTACTAATTTTCAAGGCCG 2258
QY 977 CTTTCCCTCCCTCCACTGGAGAA 1001
Db 2259 CTTTCCCTCCCTCCACTGGAGAA 2283
```

## RESULT 4

AAV64991  
ID AAV64991 standard; cDNA; 2928 BP.

AC AAV64991;

XX AAV64991;

DT 05-FEB-1999 (first entry)

XX Human RAR-alpha cDNA.

DE Fusion protein; myl; retinoic acid receptor-alpha; RAR; human;

XX acute promyelocytic leukaemia; APL; t(15;17); translocation;

KW treatment; all-trans retinoic acid; ss.

KW Homo sapiens.

OS Homo sapiens.

XX Key Location/Qualifiers

FT 103..1491

FT /\*tag= a

FT /product= "RAR-alpha"

/note= "retinoic acid receptor-alpha"

US5843642-A.

01-DEC-1998.

21-JUL-1993; 93US-0095728.

21-JUL-1993; 93US-0095728.

22-MAR-1991; 91US-0673838.

22-MAR-1991; 91US-0675084.

(SLOK ) SLOAN KEYTERING INST CANCER RES.

Dmitrovsky E, Frankel S, Miller WH, Warrell RP;

WPI; 1999-044563/04.

P-PSDB; AAW81964.

Diagnosis of acute promyelocytic leukaemia - by detecting nucleic acid encoding abnormal retinoic acid receptor-alpha

Disclosure; Column 29-34; 38pp; English.

This sequence encodes the human retinoic acid receptor alpha, RAR-alpha which is used in a method for identifying a subject with acute promyelocytic leukaemia (APL) resulting from a t(15;17) translocation who will respond to treatment with all-trans retinoic acid. The protein can also be used to identify a subject with indications of APL who will not respond to treatment with all-trans retinoic acid.

Sequence 2928 BP; 582 A; 1011 C; 789 G; 546 T; 0 other;

Query Match 87.1%; Score 871.8; DB 20; Length 2928;

Best Local Similarity 94.4%; Pred. No. 3.2e-181;

Matches 873; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 77 CCGTNN 136

Db 1359 CTTGGACACTCTCAGCGGACAGCGGGGGTGGGGGCGGAGCGGGGTGGCTGGCC 1418

QY 137 CCGGCCAGGAGTGTAGCCCGAGCTCAGCCCGAGCTCCACAGAGAGCAGCGGCGAC 196

Db 1419 CCGGCCAGGAGTGTAGCCCGAGCTCAGCCCGAGCTCCACAGAGAGCAGCGGCGAC 1478

QY 197 CCACTCCCGTGACCGCCACGCGACATGGACAGCCCTCGCCCTCCGCCCGGCTTTT 256

Db 1479 CCACTCCCGTGACCGCCACGCGACATGGACAGCCCTCGCCCTCCGCCCGGCTTTT 1538

QY 257 CTCTGCTTTTCTACCGACCATGTGACCCCGCACAGCCCTGCCCGCCACCTGCCCTCC 316

Db 1539 CTCTGCTTTTCTACCGACCATGTGACCCCGCACAGCCCTGCCCGCCACCTGCCCTCC 1598

QY 317 GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376

Db 1599 GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658

QY 377 AGGCTGGGCGCTCAGTGGAGTGCCTGCCACAGCGCTGGGCTGAGTGAGGCGCGAG 436

Db 1659 AGGCTGGGCGCTCAGTGGAGTGCCTGCCACAGCGCTGGGCTGAGTGAGGCGCGAG 1718

QY 437 GCCAGGAAGTGAAGTGGGCGGCTGTCTGGGTCTCAGAGTGGGTCTCGGGGCGCTCGTG 496

Db 1719 GCCAGGAAGTGAAGTGGGCGGCTGTCTGGGTCTCAGAGTGGGTCTCGGGGCGCTCGTG 1778

QY 497 TTCATCAAGACACCCCTGTGCCAGCTCAGCAGCATTTTATCAGCAGCAAGCCAGGAC 556

Db 1779 TTCATCAAGACACCCCTGTGCCAGCTCAGCAGCATTTTATCAGCAGCAAGCCAGGAC 1838

QY 557 TTGGCTCCCCATCCTCAGAACTCAAGCCATTGCTCCCGACAGTGGGGACCTCAACCT 616

Db 1839 TTGGCTCCCCATCCTCAGAACTCAAGCCATTGCTCCCGACAGTGGGGACCTCAACCT 1898

QY 617 CCCCCCTCCCTCGTGTGTGACAGAGGGGGTGGGACAGGGGGGGGGTCCCTCGTAC 676  
DB 1899 CCCCCCTCCCTCGTGTGTGACAGAGGGGGTGGGACAGGGGGGGGGTCCCTCGTAC 1958  
QY 677 ATACCCTGCCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTATTTAAATTT 736  
DB 1959 ATACCCTGCCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTATTTAAATTT 2018  
QY 737 TTTTGTGTTGATTTTTTAATAAGAAATTTTCATTTTAAGCACATTTATACGAAGAATT 796  
DB 2019 TTTTGTGTTGATTTTTTAATAAGAAATTTTCATTTTAAGCACATTTATACGAAGAATT 2078  
QY 797 TGTGCTGTGTATTTGGGGAGCTGGATCCAGAGCTGGAGGGGGTGGTCCGGGGAGGGA 856  
DB 2079 TGTGCTGTGTATTTGGGGAGCTGGATCCAGAGCTGGAGGGGGTGGTCCGGGGAGGGA 2138  
QY 857 GTGCTCGGAAGGGGGCCCCACTCTCTTTCATCTCCCTGTCCTGCTGCTGCTGCT 916  
DB 2139 GTGCTCGGAAGGGGGCCCCACTCTCTTTCATCTCCCTGTCCTGCTGCTGCTGCT 2198  
QY 917 CAGCCTTTTCTCTCAGTTTTTCTTTTAAACTGTGAAGTACTAACTTTTCAAGGCTG 976  
DB 2199 CAGCCTTTTCTCTCAGTTTTTCTTTTAAACTGTGAAGTACTAACTTTTCAAGGCTG 2258  
QY 977 CTTTCCCTCCCTCCACTGGAGAA 1001  
DB 2259 CTTTCCCTCCCTCCACTGGAGAA 2283

## RESULT 5

AAN90124

ID AAN90124 standard; DNA; 2940 BP.

XX

AAN90124;

XX

01-NOV-1989 (first entry)

XX

DNA of clone pHR1.

XX

Clone pHR1; DNA; retinoic acid receptor; ligand  
complexes; human.

XX

Homo sapiens (Human).

XX

Key Location/Qualifiers

FH

CDS 103..149

FT

/\*tag= a

XX

W08905355-A.

PN

15-JUN-1989.

XX

01-DEC-1988; 88WO-US04284.

XX

(SALK ) SALK INST FOR BIOLOGICAL STUD.

XX

Evans RM, Giguere V, Ong ES, Segal PS;

XX

WPI; 1989-192701/26.

DR

P-PSDB; AAP90395.

XX

DNA encoding retinoic acid receptor proteins

PT

- used to produce proteins for studying complexes with

PT

ligands and in diagnostic assays.

XX

Disclosure; fig. 1B; 75pp; English.

XX

DNA of clone pHR1 contg. the primary sequence of a

CC

protein (see corresp. AAP90395) that has ligand binding and

CC

transcription activating properties of retinoic acid receptor

CC

(RAR) protein. Used to make chimeric receptors,

CC to produce receptor, to study binding complexes, and to screen cpds.  
CC for RAR-agonists and antagonists.

XX  
SQ Sequence 2940 BP; 594 A; 1010 C; 790 G; 546 T; 0 other;

Query Match 87.1%; Score 871.8; DB 10; Length 2940;  
Best Local Similarity 94.4%; Pred. No. 3.2e-181;  
Matches 873; Conservative 0; Mismatches 52; Indels 0; Gaps 0;

QY 77 COTNN 136  
DB 1359 CCGTGACACTCTGAGCGACAGCCGGGGGTGGGGCGGAGCGGGGTGGCCCTGGCCCC 1418  
QY 137 CCGCCAGGACGCTCTAGCCCCAGCCTCAGCCCCAGCTCCACAGAGAGAGCCCGGCGAC 196  
DB 1419 CCGCCAGGACGCTCTAGCCCCAGCCTCAGCCCCAGCTCCACAGAGAGAGCCCGGCGAC 1478  
QY 137 CCACCTCCCGTACCGCCGACCCACATGACACAGCCCTCGCCCTCCCGCCCGGCTTTT 256  
DB 1479 CCACCTCCCGTACCGCCGACCCACATGACACAGCCCTCGCCCTCCCGCCCGGCTTTT 1538  
QY 257 CTCGCTCTTCTACCGACATGTGACCCGACAGCCCTGCCACCTGCCCTCCCGCTCCCG 316  
DB 1539 CTCGCTCTTCTACCGACATGTGACCCGACAGCCCTGCCACCTGCCCTCCCGCTCCCG 1598  
QY 317 GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376  
DB 1599 GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658  
QY 377 AGGCTGGGGCTCAGTGGACTGCTGCTCCACAGCTGGGCTGACGCTCAGGAGGAGGAG 436  
DB 1659 AGGCTGGGGCTCAGTGGACTGCTGCTCCACAGCTGGGCTGACGCTCAGGAGGAGGAG 1718  
QY 437 GCCAGGAATGAGTGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 496  
DB 1719 GCCAGGAATGAGTGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1778  
QY 497 TTCATCAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAAAAGCCAGGAC 556  
DB 1779 TTCATCAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAAAAGCCAGGAC 1838  
QY 557 TTGGCTCCCTCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGAGGAGGAGGAGGAG 616  
DB 1839 TTGGCTCCCTCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGAGGAGGAGGAG 1898  
QY 617 CCCCCCTGCTCGGTGGTGACAGAGGGGGTGGGACAGGGGGGGGGTTCCTCCCTGTAC 676  
DB 1899 CCCCCCTGCTCGGTGGTGACAGAGGGGGTGGGACAGGGGGGGGGTTCCTCCCTGTAC 1958  
QY 577 ATACCCTGCCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTATTTAAATTT 736  
DB 1959 ATACCCTGCCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTATTTAAATTT 2018  
QY 737 TTTTGTGTTGATTTTTTAATAAGAAATTTTCATTTTAAGCACATTTATACGAAGAATT 796  
DB 2019 TTTTGTGTTGATTTTTTAATAAGAAATTTTCATTTTAAGCACATTTATACGAAGAATT 2078  
QY 797 TGTGCTGTGTATTTGGGGAGCTGGATCCAGAGCTGGAGGGGGTGGGTCGGGGGAGGGA 856  
DB 2079 TGTGCTGTGTATTTGGGGAGCTGGATCCAGAGCTGGAGGGGGTGGGTCGGGGGAGGGA 2138  
QY 857 GTGCTCGGAAGGGGGCCCCACTCTCTTTCATCTCCCTGTCCTGCTGCTGCTGCTGCT 916  
DB 2139 GTGCTCGGAAGGGGGCCCCACTCTCTTTCATCTCCCTGTCCTGCTGCTGCTGCTGCT 2198  
QY 917 CAGCCTTTTCTCTCAGTTTTTCTTTTAAACTGTGAAGTACTAACTTTTCAAGGCTG 976  
DB 2199 CAGCCTTTTCTCTCAGTTTTTCTTTTAAACTGTGAAGTACTAACTTTTCAAGGCTG 2258  
QY 977 CTTTCCCTCCCTCCACTGGAGAA 1001  
DB 2259 CTTTCCCTCCCTCCACTGGAGAA 2283











Db 1728 GCCAGGAACGAGTGAAGCCCTGGTCTGGGTCTCAGGATGGTCTGGGGGCTCGTG 1787  
QY 497 TTATCAAGACACCCCTCTGCCAGCTCACACACATCTTCATCACAGCAAAACGCCAGGAC 556  
Db 1788 TTATCAAGACACCCCTCTGCCAGCTCACACACATCTTCATCACAGCAAAACGCCAGGAC 1847  
QY 557 TTGGCTCCCCCAGCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAACCTCAACT 616  
Db 1848 TTGGCTCCCCCAGCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAACCTCAACT 1907  
QY 617 CCCCCCTGCCTGGTGTGACAGAGGGGTGGACAGGGGGTCCCGCTGTAC 676  
Db 1908 CCCCCCTGCCTGGTGTGACAGAGGGGTGGACAGGGGGTCCCGCTGTAC 1967  
QY 677 ATACCTCGCCATACCAACCCAGGTATTAATCTCGTGGTGTGTTTTTATTTAATTT 736  
Db 1968 ATACCTCGCCATACCAACCCAGGTATTAATCTCGTGGTGTGTTTTTATTTAATTT 2027  
QY 737 TTTGTTTTGATTTTTTAAGAATTTTCATTTTAAGCACATTTATACGAAGAA 794  
Db 2028 TTTGTTTTGATTTTTTAAGAATTTTCATTTTAAGCACATTTATTTAAGCAAAAAA 2085

RESULT 11  
AAQ29334.  
ID AAQ29334 standard; cDNA; 3036 BP.  
XX  
AC AAQ29334;  
XX  
DT 09-MAR-1993 (first entry)  
XX  
DE myl/RAR-alpha fusion.  
XX  
KW Retinoic acid receptor; RAR-alpha; myl; acute promyelocytic leukemia;  
APb; translocation; chromosome 17; chromosome 15; PCR; primer;  
KW [t(15;17)(q21;q11-22)]; breakpoint; polymerase chain reaction; ss.  
XX  
OS Synthetic.

Key Location/Qualifiers  
FT CDS 67..2457  
FT /\*tag= a  
FT primer\_bind 715..744  
FT /\*tag= b  
FT primer\_bind 976..1001  
FT /\*tag= c  
FT primer\_bind 1382..1413  
FT /\*tag= d  
XX  
XX WO9216660-A.  
XX  
XX 01-OCT-1992.  
XX  
XX 23-MAR-1992; 92WO-US02320.  
XX  
XX 22-MAR-1991; 91US-0673838.  
XX  
XX 22-MAR-1991; 91US-0675084.  
XX  
XX (SLOK ) SLOAN KETTERING INST CANCER.  
XX  
XX Dmitrovsky E, Evans RM, Frankel S, Kazizuka A, Miller WH;  
PI Warrell RP;  
XX  
XX WPI; 1992-349240/42.  
XX  
XX P-PSDB; AAR27533.  
XX  
XX Marker for acute promyelocytic leukaemia and other neoplasias -  
PT comprising nucleic acid and encoded abnormal retinoic acid  
PT receptor-alpha receptor  
XX  
XX Disclosure; Fig 4; 84pp; English.  
XX  
XX

CC The sequence given shows a fusion between retinoic acid receptor  
CC (RAR)-alpha and myl which is characteristic of acute promyelocytic  
CC leukemia (APL). This is caused by a translocation of a portion of  
CC the long arm of chromosome 17 onto the long arm of chromosome 15  
CC [t(15;17)(q21;q11-22)]. The breakpoint region has been cloned and it  
CC has been shown that DNA rearrangements are clustered in the region  
CC of the first intron of RAR-alpha. This sequence was isolated by  
CC polymerase chain reaction (PCR) using primers which correspond to  
CC sequences both 5' and 3' to the breakpoint region.  
XX  
SQ Sequence 3036 BP; 605 A; 1018 C; 890 G; 523 T; 0 other;  
Query Match 64.8%; Score 648.2; DB 13; Length 3036;  
Best Local Similarity 92.5%; Pred. No. 2.9e-132; Indels 0; Gaps 0;  
Matches 650; Conservative 0; Mismatches 53;  
QY 77 CGTNNNGGCCCC 136  
Db 2328 CTTGACACTCTGAGGGACAGCCGGGGGTGGGGGGGGGGGGTGGCTGGCCCC 2387  
QY 137 CCGCCAGGAGCTGTAGCCCCAGCCTCAGCCCCAGCTCCAAAGAGAGAGCCCGGCCAC 196  
Db 2388 CCGCCAGGAGCTGTAGCCCCAGCCTCAGCCCCAGCTCCAAAGAGAGAGCCCGGCCAC 2447  
QY 197 CCACCTCCCGCTGACCGCCCGCCACGCCACATGGACACAGCCCTCGCCCTCGCGCTTTT 256  
Db 2448 CCACCTCCCGCTGACCGCCCGCCACGCCACATGGACACAGCCCTCGCGCTTTT 2507  
QY 257 CTCTGCTTTTACCGACCATGTGACCCCGCAGCCAGCCCTGCCCCACCTCGCCCTCCCG 316  
Db 2508 CTCTGCTTTTACCGACCATGTGACCCCGCAGCCAGCCCTGCCCCACCTCGCCCTCCCG 2567  
QY 317 GCAGTACTGGGGACCTTCCTCTGGGGACGGGGAGGAGGAGGAGGAGCTCTCTTGACAG 376  
Db 2568 GCAGTACTGGGGACCTTCCTCTGGGGACGGGGAGGAGGAGGAGGAGCTCTCTTGACAG 2627  
QY 377 AGGCTTGGCCCTCAGTGTGACCTGCTGCCACAGCCCTGGGGTGTAGCTCAGAGGCCGAG 436  
Db 2628 AGGCTTGGCCCTCAGTGTGACCTGCTGCCACAGCCCTGGGGTGTAGCTCAGAGGCCGAG 2687  
QY 437 GCAGGAACGTAGTGAAGCCCTCTGGTGGTCTCAGATGGTCTCTGGGGGCTCTGTG 496  
Db 2688 GCAGGAACGTAGTGAAGCCCTCTGGTGGTCTCAGATGGTCTCTGGGGGCTCTGTG 2747  
QY 497 TTCATCAAGACACCCCTCTGCCAGCTCACACATCTTCATCACAGCAAAACGCCAGGAC 556  
Db 2748 TTCATCAAGACACCCCTCTGCCAGCTCACACATCTTCATCACAGCAAAACGCCAGGAC 2807  
QY 557 TTGGCTCCCCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGGAACCTCAACCT 616  
Db 2808 TTGGCTCCCCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGGAACCTCAACCT 2867  
QY 617 CCGCCCTGCCTGGTGTGACAGAGGGGTGGGAGGGGGGGGGGGTCCCGCTGTAC 676  
Db 2868 CCGCCCTGCCTGGTGTGACAGAGGGGTGGGAGGGGGGGGGGGTCCCGCTGTAC 2927  
QY 677 ATACCTGCGCATACCAACCCAGGTATTAATCTCGTGGTGTGTTTTTATTTAATTT 736  
Db 2928 ATACCTGCGCATACCAACCCAGGTATTAATCTCGTGGTGTGTTTTTATTTAATTT 2987  
QY 737 TTTGTTTTGATTTTTTAAAGAAATTTTCATTTTAAGCAACA 779  
Db 2988 TTTGTTTTGATTTTTTAAAGAAATTTTCATTTTAAGCAACA 3030  
RESULT 12  
AAV20474  
ID AAV20474 standard; DNA; 3036 BP.  
XX  
AC AAV20474;  
XX  
DT 17-JUN-1998 (first entry)  
XX





3043	Db	GCAGTACTGGGGACCTTCCTCGGGGACGGGAGGGAGGCAGGACTCCTCTGGGACAG	3102
377	Qy	AGGCCTGGGCCCTCAGTGTGACTGCTGTCCACAGCCTGGGCTGAGCTCAGAGGCCGAG	436
3103	Db	AGGCCTGGGCCCTCAGTGTGACTGCTGTCCACAGCCTGGGCTGAGCTCAGAGGCCGAG	3162
437	Qy	GCCAGAACTGAGTGAAGGCCCTGGTCTCTGGGTCTCAGGATGGGTCTGGGGGCTCGTG	496
3163	Db	GCCAGAACTGAGTGAAGGCCCTGGTCTCTGGGTCTCAGGATGGGTCTGGGGGCTCGTG	3222
497	Qy	TTTCATCAGACACCCCTCTGCCCCAGCTCACACATCTTTCATCACAGCAACGCCAGGAC	556
3223	Db	TTTCATCAGACACCCCTCTGCCCCAGCTCACACATCTTTCATCACAGCAACGCCAGGAC	3282
557	Qy	TTGGCTCCCCCATCTCTCAGAACTCACAGCCATTGCTCCCCAGCTGGGGAACTCAACCT	616
3283	Db	TTGGCTCCCCCATCTCTCAGAACTCACAGCCATTGCTCCCCAGCTGGGGAACTCAACCT	3342
617	Qy	CCCCCTCGCTCGGTGGTGACAGAGGGGTGGACAGGGCGGGGGTTCCTCCCTGTAC	676
3343	Db	CCCCCTCGCTCGGTGGTGACAGAGGGGTGGACAGGGCGGGGGTTCCTCCCTGTAC	3402
677	Qy	ATACCTGGCATCAACACCCCAAGTATTAATTCCTCGCTGGTGTGTTTTATTTAAATTT	736
3403	Db	ATACCTGGCATCAACACCCCAAGTATTAATTCCTCGCTGGTGTGTTTTATTTAAATTT	3462
737	Qy	TTTTGTTTTGATTTTTTTTAAATGAATTTTCATTTTAAAGCACA	779
3463	Db	TTTTGTTTTGATTTTTTTTAAATGAATTTTCATTTTAAAGCAAA	3505

RESULT	15
AAT33259	
ID	AAT33259 standard; CDNA; 3511 BP.
XX AC AAT33259;	
XX DT 09-OCT-1996	(first entry)
XX DE PML-RAR-alpha DNA sequence.	
DE Hammerhead ribozyme; acute promyelocytic leukaemia; APL;	
KW lymphoma; therapy; PML-RAR-alpha; retinoic acid receptor;	
KW external guide sequence; EGS; antisense; ss.	
XX Homo sapiens.	
OS OS	
XX Key Location/Qualifiers	
FT misc_feature 1..1724	
FT /tag= a	
/note= "PML DNA"	
FT misc_difference 1477	
FT /*tag= b	
/note= "base n at position 1477 is not identified in the specification"	
FT misc_feature 1725...3511	
FT /*tag= C	
/note= "RAR-alpha DNA"	
PN WO9618733-A2.	
XX PD 20-JUN-1996.	
XX PF 14-DEC-1995;	95WO-US16451.
XX PR 14-DEC-1994;	94US-0354956.
XX (INNO-) INNOVIR LAB INC.	
PA George ST,	Pace U;
PI Goldberger AR,	
XR WPI;	1996-300650/30.

RNA construct(s) including ribozyme(s) and antisense oligo:nucleotide(s) - for the inactivation of RNA associated with, e.g. promyelocytic leukaemia or follicular lymphoma

Example 1; Page 50-52; 81pp; English.

PML-RAR-alpha DNA (AAT33259) is associated with acute promyelocytic leukaemia (APL). APL is characterised by a balanced, reciprocal translocation between the long arms of chromosomes 15 and 17, resulting in a fusion of the retinoic acid receptor gene (RAR-alpha) and a gene for a putative transcription factor, PML. Methods for treating APL involve the use of ribozymes, external guide sequences and antisense oligonucleotides (see also AAT33245 and AAT33248-57) that specifically cleave the PML-RAR-alpha fusion mRNA (see also AAT33251) but not wild-type RAR-alpha mRNA. Plasmids carrying portions of the PML-RAR-alpha and RAR-alpha genes (see also AAT33260) allow synthesis of shortened versions of APL mRNA in vitro, facilitating testing and screening processes.

Sequence 3511 BP; 715 A; 1168 C; 1038 G; 589 T; 1 other

[illegible]

Search completed: March 30, 2003, 10:14:57  
Job time : 250.299 secs

---



GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 09:12:13 ; Search time 33.7584 Seconds  
(without alignments)  
9093.534 Million cell updates/sec

Title: 'US-09-691-220-3\_COPY\_15000\_16000

Perfect score: 1001

Sequence: 1 gctaaaaattccctgtt.....ccctccctcccaactggagaa 1001

Scoring table: IDENTITY\_NVC

Gapop 10.0 , Capext 1.0

Searched: 441362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 892724

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

: Listing first 45 summaries

Database : Issued Patents\_NA:\*

1: /cgn2\_6/ptodata/1/ina/5A\_COMB.seq:\*

2: /cgn2\_6/ptodata/1/ina/5B\_COMB.seq:\*

3: /cgn2\_6/ptodata/1/ina/6A\_COMB.seq:\*

4: /cgn2\_6/ptodata/1/ina/6B\_COMB.seq:\*

5: /cgn2\_6/ptodata/1/ina/PCTUS\_COMB.seq:\*

6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	871.8	87.1	2928	US-08-095-728B-3	Sequence 3, Appli
2	871.8	87.1	2940	PCT-US92-02320A-3	Sequence 3, Appli
3	871.8	87.1	2940	US-08-592-383-1	Sequence 1, Appli
4	871.8	87.1	2940	5171671-1	Patent No. 5171671
5	689	68.8	2658	US-08-592-383-3	Sequence 3, Appli
6	648.2	64.8	3036	US-08-306-691B-52	Sequence 52, Appli
7	648.2	64.8	3036	US-08-095-728B-1	Sequence 1, Appli
8	648.2	64.8	3036	PCT-US92-02320A-1	Sequence 1, Appli
9	648.2	64.8	3511	US-08-892-747-13	Sequence 13, Appli
10	76	7.6	93	US-08-592-383-8	Sequence 8, Appli
11	52.4	5.2	320	US-09-165-264-11	Sequence 11, Appli
12	50.2	5.0	319	US-09-165-264-8	Sequence 8, Appli
13	49.6	5.0	320	US-09-165-264-13	Sequence 13, Appli
14	49.4	4.9	320	US-09-165-264-7	Sequence 7, Appli
15	49.4	4.9	320	US-09-165-264-14	Sequence 14, Appli
16	49	4.9	318	US-09-165-264-12	Sequence 12, Appli
17	48.8	4.9	152331	US-09-128-155-16	Sequence 16, Appli
18	48.4	4.8	7218	US-08-232-463-14	Sequence 14, Appli
19	43.4	4.3	4480	US-09-191-171-7	Sequence 7, Appli
20	43.4	4.3	4480	US-09-385-707-7	Sequence 7, Appli
21	43.2	4.3	2269	US-09-394-645-1	Sequence 1, Appli
22	43.2	4.3	2269	US-09-243-560B-1	Sequence 1, Appli
23	43	4.3	198	US-08-330-108-16	Sequence 16, Appli
24	43	4.3	198	PCT-US92-10087-16	Sequence 16, Appli
25	43	4.3	72604	US-09-268-992-7	Sequence 7, Appli
26	43	4.3	72604	US-09-657-474-7	Sequence 7, Appli
27	42.8	4.3	53526	US-08-658-136-2	Sequence 2, Appli

28 42.8 4.3 53577 3 US-08-658-136-1 Sequence 1, Appli  
29 41.8 4.2 936 4 US-09-179-558-62 Sequence 62, Appli  
30 41.6 4.2 1931 3 US-09-019-942-2 Sequence 2, Appli  
31 41.6 4.2 1931 4 US-09-099-041A-1 Sequence 1, Appli  
32 41.6 4.2 1931 4 US-09-245-281-1 Sequence 1, Appli  
33 41.6 4.2 1931 4 US-09-470-271-2 Sequence 2, Appli  
34 41.6 4.2 1931 4 US-09-207-359B-1 Sequence 1, Appli  
35 41.4 4.1 19124 2 US-08-487-826B-13 Sequence 13, Appli  
36 41.2 4.1 1553 4 US-09-280-116-10 Sequence 10, Appli  
37 41 4.1 1817 1 US-08-473-981A-5 Sequence 5, Appli  
38 41 4.1 1817 2 US-08-474-087-5 Sequence 5, Appli  
39 41 4.1 3602 4 US-09-402-929-1 Sequence 1, Appli  
40 41 4.1 7808 2 US-08-149-097D-22 Sequence 22, Appli  
41 41 4.1 7808 3 US-08-949-386-22 Sequence 22, Appli  
42 41 4.1 7808 3 US-08-450-562-22 Sequence 22, Appli  
43 41 4.1 7808 4 US-08-984-709A-22 Sequence 22, Appli  
44 41 4.1 7808 4 US-08-450-272-22 Sequence 22, Appli  
45 40.6 4.1 1364 1 US-08-265-087-3 Sequence 3, Appli

## ALIGNMENTS

## RESULT 1

US-08-095-728B-3

; Sequence 3, Application US/08095728B

; Patent No. 5843642

; GENERAL INFORMATION:

; APPLICANT: DMITROVSKY, ETHAN

; APPLICANT: WARRELL JR, RAYMOND P

; APPLICANT: MILLER JR, WILSON H

; APPLICANT: FRANKEL, STANLEY

; TITLE OF INVENTION: METHODS FOR THE DETECTION AND

; NUMBER OF SEQUENCES: 11 TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA (APL)

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: COOPER & DUNHAM LLP

; STREET: 1185 AVENUE OF THE AMERICAS

; CITY: NEW YORK

; STATE: NEW YORK

; COUNTRY: U.S.A.

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; OPERATING SYSTEM: IBM PC compatible

; SOFTWARE: PatentIn Release #1.24

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/095.728B

; FILING DATE: 21-JUL-1993

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/673,838

; FILING DATE: 22-MAR-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: WHITE, JOHN P

; REGISTRATION NUMBER: 28,678

; REFERENCE/DOCKET NUMBER: 38694-A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (212) 278-0400

; TELEFAX: (212) 391-0525

; INFORMATION FOR SEQ ID NO: 3:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2928 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

; IMMEDIATE SOURCE:

; CLONE: hRAR ALPHA

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 103..1488





Db 2079 TGTGCTGTGTATTGGGGGAGCTGGATCCAGAGCTGGAGGGGTGGGTCCGGGGAGGGA 2138  
QY 857 GTGGCTCGGAAGGGGCGCCACACTCTCTTCATGCTCCCTGTGCCCCCAGTTCTCCCT 916  
Db 2139 GTGGCTCGGAAGGGGCGCCACACTCTCTTCATGCTCCCTGTGCCCCCAGTTCTCCCT 2198  
QY 917 CAGCCCTTTTCCCTCCTCAGTTTCTCTTTAAACTGTGAAGTACTAACTTTCCCAAGGCTG 976  
Db 2199 CAGCCCTTTTCCCTCCTCAGTTTCTCTTTAAACTGTGAAGTACTAACTTTCCCAAGGCTG 2258  
QY 977 CTTTCCCTCCCTCCCACTGGAGAA 1001  
Db 2259 CCTTCCCTCCCTCCCACTGGAGAA 2283  
RESULT 4  
5171671-1  
; Patent No. 5171671  
; APPLICANT: EVANS, RONALD M.; ONG, ESTELITA S.; SEGUI,  
; PRUDIMAR S.; THOMPSON, CATHERINE C.; UEMSONO, KAZUHIKO  
; GUGIERE, VINCENT  
; TITLE OF INVENTION: RETINOIC ACID RECEPTOR COMPOSITION  
; NUMBER OF SEQUENCES: 2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/546,256  
; FILING DATE: 06-AUG-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 276,536  
; FILING DATE: 30-NOV-1988  
; APPLICATION NUMBER: 128,331  
; FILING DATE: 02-DEC-1987  
; SEQ ID NO: 1:  
; LENGTH: 2940  
5171671-1

Query Match 87.1%; Score 871.8; DB 6; Length 2940;  
Best Local Similarity 94.4%; Pred. No. 3.3e-208;  
Matches 873; Conservative 0; Mismatches 52; Indels 0; Gaps 0;  
QY 77 CGTNN 136  
Db 1359 CCTGGACACTCTGAGCGGACAGCGGGGGGTGGGGGGGAGCGGGGTGGCCCTGGCCCC 1418  
QY 137 CCGCGCAGCAGCTGTAGCCGCCAGCTCAGCCCCAGCTCCACAGAGAGCAGCGGCCAC 196  
Db 1419 CCGCGCAGCAGCTGTAGCCGCCAGCTCAGCCCCAGCTCCACAGAGAGCAGCGGCCAC 1478  
QY 197 CCACTCCCGGTGACCGCCACGCCACATGACACAGCCCTCGCCCTCGCCCGGCTTTT 256  
Db 1479 CCACTCCCGGTGACCGCCACGCCACATGACACAGCCCTCGCCCTCGCCCGGCTTTT 1538  
QY 257 CTCTGCTTTTCTACCGACATGTGACCGCGCAGCCAGCCCTGCCCCACCTGCCCCCGG 316  
Db 1539 CTCTGCTTTTCTACCGACATGTGACCGCGCAGCCCTCGCCCTCGCCCGGCTTTT 1598  
QY 317 CGAGTACTGGGACCTTCCTGGGGAGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376  
Db 1599 CGAGTACTGGGACCTTCCTGGGGAGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658  
QY 377 AGGCTGGGCGCTCAGTGGACTGCTGCTCCACAGCCTGGGTGAGCTGAGAGGCGGAG 436  
Db 1659 AGGCTGGGCGCTCAGTGGACTGCTGCTCCACAGCCTGGGTGAGCTGAGAGGCGGAG 1718  
QY 437 GCAGGAATGAGTAGGCGCCCTGTGCTGGGTCTCAGTAGGCTCCTGGGGCGCTGGTG 496  
Db 1719 GCAGGAATGAGTAGGCGCCCTGTGCTGGGTCTCAGTAGGCTCCTGGGGCGCTGGTG 1778  
QY 497 TTATCAACACACCCCTCTGCCAGCTCAGCATCTTATCACCAGCAACGCCAGGAC 556  
Db 1779 TTATCAACACACCCCTCTGCCAGCTCAGCATCTTATCACCAGCAACGCCAGGAC 1838  
QY 557 TTGGCTCCCGCTCCTCAGAACTCACAAGCATTGCTCCCGCTGGGGAACTCAACCT 616

Db 1839 TTGCTCCCGCATCTCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAACCTCAACCT 1898  
QY 617 CCCCCCTGCGCTCGTGGTGACAGAGGGGTGGACAGGGGCGGGGGGTTCGCCCTGTAC 676  
Db 1899 CCCCCCTGCGCTCGTGGTGACAGAGGGGTGGACAGGGGCGGGGGGTTCGCCCTGTAC 1958  
QY 677 ATACCTCGCATACCAACCCAGGTATTAATCTCGCTGGTGTGTTTTTATTTAAATTT 736  
Db 1959 ATACCTCGCATACCAACCCAGGTATTAATCTCGCTGGTGTGTTTTTATTTAAATTT 2018  
QY 737 TTTTGTGTTGATTTTTTAAAGAAATTTTCAATTTAAGCACATTTATPACTGAAGAATT 796  
Db 2019 TTTTGTGTTGATTTTTTAAAGAAATTTTCAATTTAAGCACATTTATPACTGAAGAATT 2078  
QY 797 TGTGCTGTGATTGGGGGAGCTGGATCCAGAGCTGGAGGGGTGGGTCCGGGGGAGGGA 856  
Db 2079 TGTGCTGTGATTGGGGGAGCTGGATCCAGAGCTGGAGGGGTGGGTCCGGGGGAGGGA 2138  
QY 857 GTGGCTCGGAAGGGGCGCCCACTCTCTTTCATGCTCCCTGTGCCCGCCAGTTCTCTCT 916  
Db 2139 GTGGCTCGGAAGGGGCGCCCACTCTCTTTCATGCTCCCTGTGCCCGCCAGTTCTCTCT 2198  
QY 917 CAGCCTTTTCTCTCAGTTTCTCTTTAAAACTGTGAAGTACTAACTTTCAAGGCGCTG 976  
Db 2199 CAGCCTTTTCTCTCAGTTTCTCTTTAAAACTGTGAAGTACTAACTTTCAAGGCGCTG 2258  
QY 977 CCTTCCCTCCCTCCCACTGGAGAA 1001  
Db 2259 CCTTCCCTCCCTCCCACTGGAGAA 2283  
RESULT 5  
US-08-592-383-3  
; Sequence 3, Application US/08592383  
; Patent No. 5830760  
; GENERAL INFORMATION:  
; APPLICANT: Tsal, S. and S.J. Collins  
; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acid  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage  
; COMPUTER: IBM PC/386 Compatible  
; OPERATING SYSTEM: MS-DOS 4.01  
; SOFTWARE: Word for Windows 5.01-t  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/592,383  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/099,242  
; FILING DATE: July 28, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Broderick, Thomas F.  
; REGISTRATION NUMBER: 31,332  
; REFERENCE/DOCKET NUMBER: FPCR-1-7190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2658 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; DESCRIPTION: page 4, RAR-alpha403 dominant negative; deleted of 1311-1596  
; IMMEDIATE SOURCE:

LIBRARY: CDNA  
US-08-592-383-3

Query Match : 68.8%; Score 689; DB 2; Length 2658;  
Best Local Similarity 100.0%; Pred. No. 1.3e-162;  
Matches 689; Conservative 0; Mismatches 0; Indels

[illegible]

Qy	373	ACAGAGGCTGGGCCCTCAGTGGACTGCCTCCACAGCTGGGTGACGTCAGAGGC	432
Db	1373	ACAGAGGCTGGGCCCTCAGTGGACTGCCTCCACAGCTGGGTGACGTCAGAGGC	1432

Qy 433 CGAGGCCAGGAAC TGAGTCAGGGCCCCCTGGTCTGGGTCTCAGGATGGGTCTCTGGGGGGCT 492  
 |||||  
 Db 1433 CGAGGCCAGGAAC TGAGTCAGGGCCCCCTGGTCTGGGTCTCAGGATGGGTCTCTGGGGGGCT 1492

Qy 493 CGTGTTCATCAACACACACCCCTCTGCCCGAGTCCACACATCTTTCATCACCAGCAACGCCA 552  
|||||  
Db 1493 CGTGTTCATCAACACACCCCTCTGCCCGAGTCCACACATCTTTCATCACCAGCAACGCCA 1552

Qy 553 GGAATTGGCTCCCCCATCTCTCAGAACTCTCAGAGCCATTGCTCCCCAGCTGGGGAACTCA 612

Dh 1553 GGACTTGGCTCCCCCATCTCTCAGAACTCTCAGAGCCATTGCTCCCCAGCTGGGGAACTCA 1612

[illegible]

673 GTACATACCCTGCCATACCAACCCAGGATTAATCTCGCTGGTTTGTTTATTTA 732

QY 733 A T T T T T T T G T T T T G A T T T T T T T T A A T A A G A A T T T C A T T T T A A G C A C A T T T A T A C T G A A G G 792

Qy 793 AATTTGTCTGTCTATTGGGGGGAGCTGGATCCAGAGCTGGAGGGGTGGTCCGGGGA 852

QY 853 GGGAGTGGCTCGGAAGGGGCCCCCACTCTCCTTTTCATGTCCCTGTGCCCCCAGTTCTCC 912

QY 913 TCCTCAGCGCTTTTCCTCCTCAGTTTTCTCTTTAAACTGTGAAGTACTACTTTCACGG 972

Qy 973 CCGCCTTCCCTCCCCTCCACTGGAGAA 1001  
|||||  
|||||

RESULT 6

; Sequence 52, Application US/08306691B  
 ; Patent No. 5734039  
 ; GENERAL INFORMATION:

APPLICANT: Calabrella, Bruno  
APPLICANT: Skorski, Tomasz  
TITLE OF INVENTION: ANTISENSE  
TITLE OF INVENTION: OLIGONUCLEOTIDES TARGETING COOPERATING ONCOGENES

NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Seidel, Gonda, Lavorgna & Monaco, P.C.  
STREET: Two Penn Center, Suite 1800

CITY: Philadelphia  
STATE: Pennsylvania  
COUNTRY: U.S.A.  
ZIP: 19102

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 Inch,  
COMPUTER: IBM PS/2  
OPERATING SYSTEM: MS-DOS  
SOFTWARE: WordPerfect 5.1  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/306,691B  
FILING DATE: September 15, 1994  
CLASSIFICATION: 514  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:--  
NAME: Monaco, Daniel A.  
REGISTRATION NUMBER: 30,480  
REFERENCE/DOCKET NUMBER: 8321-8  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 568-8383  
TELEFAX: (215) 568-5549  
TELEX: No. 5734039e  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3036 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-306-691B-52

[illegible]

## RESULT 6

US-08-306-691B-52  
; Sequence 52, Application US/08306691B  
; Patent No. 5734039

GENERAL INFORMATION:  
APPLICANT: Calabretta, Bruno  
APPLICANT: Skorski, Tomasz

; TITLE OF INVENTION: ANI  
 ; TITLE OF INVENTION: OLI  
 ; NUMBER OF SEQUENCES: 55  
 ; CORRESPONDENCE ADDRESS:

ADDRESSEE: Seidel, Go  
STREET: Two Penn Cent  
CITY: Philadelphia

```

; STATE: Pennsylvania
; COUNTRY: U.S.A.
; ZIP: 19102

```













US-09-165-264-14

```
Query Match.      4.9%; Score 49.4; DB 4; Length 320;
Best Local Similarity 42.0%; Pred. No. 0.0019;
Matches 113; Conservative 0; Mismatches 156; Indels 0; Gaps 0;

QY 46 CTTGGGTGTTAATCCGAATGGGGGCCCCCGTNNNNNNNNNNNNNNNNNNNN 105
Db   || ||| || | |||
QY 106 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 165
Db   || ||| || | |||
QY 166 GCCCCAGCTCAACAGACAGCCCGGCGCACCCACTCCCGTGACCGCCAGCGCACATG 225
Db   || || | || | ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 199 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 140
Db   || || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 226 GACAGAGCCCTCGGCGCTCGGCGCGGGCTTTCTGCGCTTTCTACCGACCATGTGACCCC 285
Db   || || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 286 GCACGAGCCCTGCGCCGACCTGCCCTCCC 314
Db   || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 79 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 51
Db   || || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: March 30, 2003, 13:57:00  
Job time : 59.7584 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 13:12:08 ; Search time 75.7689 Seconds  
(without alignments)  
11242.392 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_15000\_16000  
Perfect score: 1001  
Sequence: 1 gcctaaatcccggtt.....ccctccctccactggagaa 1001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 574371 seqs, 425486471 residues

Total number of hits satisfying chosen parameters: 1148742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	871.8	87.1	2907	10	US-09-954-456-524
C 2	578	57.7	595	10	US-09-822-830A-614
C 3	54.4	5.4	671	9	US-10-184-644-346
C 4	52.2	5.2	12733	9	US-10-032-393-47
C 5	52.2	5.2	12739	9	US-10-032-393-8
C 6	48.8	4.9	152331	9	US-10-035-407-16
C 7	48.2	4.8	934	9	US-10-078-090-83
C 8	46	4.6	569	9	US-10-012-542-44
C 9	45.4	4.5	1064	10	US-09-804-682-29
C 10	44.4	4.4	266	10	US-09-960-352-3882
C 11	44.4	4.4	1333	9	US-10-102-806-28
C 12	44	4.4	227	10	US-09-960-352-14543
C 13	44	4.4	337	10	US-09-960-352-13784
C 14	44	4.4	689	9	US-10-001-887-14
C 15	44	4.4	1545	10	US-09-960-301-392
C 16	43.6	4.4	328	10	US-09-960-352-14268
C 17	43.4	4.3	7428	9	US-10-147-026-11
C 18	43.2	4.3	287	10	US-09-823-101-9
C 19	43.2	4.3	1014	9	US-10-102-806-305

C	20	43.2	4.3	2269	10	US-09-860-298-1	Sequence 1, Appli
C	21	43	4.3	272	10	US-09-960-352-6986	Sequence 6986, Ap
C	22	43	4.3	393	10	US-09-960-352-4582	Sequence 4582, Ap
C	23	43	4.3	748	10	US-09-910-943-361	Sequence 361, Appl
C	24	43	4.3	1468	10	US-09-764-870-40	Sequence 40, Appl
C	25	43	4.3	1998	10	US-09-764-903-19	Sequence 19, Appl
C	26	42.8	4.3	1065	10	US-09-804-682-33	Sequence 33, Appl
C	27	42.8	4.3	2756	10	US-09-925-301-351	Sequence 351, Appl
C	28	42.8	4.3	53522	9	US-09-904-968A-1	Sequence 1, Appli
C	29	42.4	4.2	694	10	US-09-846-258-3	Sequence 3, Appli
C	30	42.2	4.2	259	10	US-09-960-352-3453	Sequence 3453, Ap
C	31	42.2	4.2	259	10	US-09-960-352-7233	Sequence 7233, Ap
C	32	42.2	4.2	869	10	US-09-925-301-359	Sequence 359, App
C	33	42.2	4.2	3372	10	US-09-888-615-18	Sequence 18, Appl
C	34	42	4.2	1923	9	US-10-042-141-12	Sequence 12, Appl
C	35	42	4.2	1923	10	US-09-726-643-12	Sequence 12, Appl
C	36	42	4.2	43058	10	US-09-954-456-292	Sequence 292, App
C	37	42	4.2	43058	10	US-09-954-456-529	Sequence 529, App
C	38	42	4.2	43058	10	US-09-880-107-3950	Sequence 3950, Ap
C	39	41.8	4.2	347	10	US-09-960-352-3371	Sequence 3371, Ap
C	40	41.8	4.2	373	10	US-09-960-352-3655	Sequence 3655, Ap
C	41	41.8	4.2	413	10	US-09-960-352-2919	Sequence 2919, Ap
C	42	41.8	4.2	487	10	US-09-960-352-7373	Sequence 7373, Ap
C	43	41.8	4.2	713	9	US-10-091-483-136	Sequence 136, App
C	44	41.8	4.2	713	10	US-09-764-846-136	Sequence 136, App
C	45	41.8	4.2	936	10	US-09-974-300-836	Sequence 836, App

ALIGNMENTS

RESULT 1

US-09-954-456-524  
; Sequence 524, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Usi  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 524  
; LENGTH: 2907  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-456-524

Query Match 87.1%; Score 871.8; DB 10; Length 2907;  
Best Local Similarity 94.4%; Pred. No. 2.9e-211;  
Matches 873; Conservative 0; Mismatches 52; Indels 0; Gaps 0;



```
; APPLICANT: Desnoyers, Luc
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Pan, James
; APPLICANT: Smith, Victoria
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3430RIC227
; CURRENT APPLICATION NUMBER: US/10/184,644
; CURRENT FILING DATE: 2002-06-28
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 612
; SEQ ID NO 346
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo Sapient
US-10-184-644-346
```

```
Query Match 5.4%; Score 54.4; DB 9; Length 671;
Best Local Similarity 6.9%; Pred. No. 0.00033;
Matches 37; Conservative 182; Mismatches 316; Indels 0; Gaps 0;
```

```
QY 23 CACCACCTGCTGAAACCCCAAGCTGGGTGTTAATCCGAATTTGGGGGGCCCCCGCTNNN 82
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 563 MNB.M.BT..M.C.BISM.YARB.TTM.KBHMNTMYI.MM.ATBBHANM....SAB..AC 504
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 83 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 142
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 503 ....TN.GYMBH.SBHM.T.M..SD.M..M..BS...MM.BMYHKS.C.SYMTAY.BMB.. 444
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 143 AGGAGCTGTAGCCAGCTCAGCCAGCTCAGCCAGCTCAGCCAGCTCAGCCAGCTCAGCTC 202
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 443 MT..M..YBYN.M.C.MYT..ASSS..MNTSMNT.Y..T...SHSS..CYCYCMYCY 384
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 203 CCCGTGACCGCCGACGATGACAGCCCTCGCCCTCGCCCTCGCCCTCGCCCTCGCCCTCG 262
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 383 MMB..HH.Y..HCSSCCSCY.T.CHTY.YYY.M..M..Y..Y..Y..Y..Y..Y..Y..Y 324
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 263 CTTTCTACCGACCATGTACCCCGACGACGACGACGACGACGACGACGACGACGACGACG 322
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 323 YT.HVYMM.S.YHB.HSHSHSSSS..Y..M.MYCY.M..M.T.MYCY..M.MBSBHS 264
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 323 CTGGGGACCTTCCTCGGGGACGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 382
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 263 SSSSSSSSTSYTKTB...MCHSHTMSHSHSTS.TMMMYCC..CYI.TYBTMM..A.. 204
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 383 GGGCCCTCAGGAGCTGCTCTCCACAGCTGGGCTGACGTCAGAGCCGAGCCAGGAGG 442
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 203 H.HSAM.S.SSS.SN...S.SBST.H.HSSTMYTMSBKM.T.AMYM.CSNHSMHSHS 144
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 443 AACTGAGTGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 502
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 143 HS.KYHSTTATATBYKBYCH.HH.HTHSCHT.NT..T.SHSSB...TSKSTD.NN.. 84
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 503 AAGACACCCCTCTCCAGCTCACCACATCTCATCATCAACCAACGACGAGCT 557
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 83 .W.C.N.CMYMN.MCRMHMGHR...HM.C.TA.DAC.....RMN...MBTCHT 29
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 4

US-10-032-393-47/c

; Sequence 47, Application US/10032393  
; Publication No. US20030027286A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Wall, Daniel

; APPLICANT: Gross, Molly

; TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE

; FILE REFERENCE: ELITRA.010A

; CURRENT APPLICATION NUMBER: US/10/032,393  
; CURRENT FILING DATE: 2001-12-21  
; PRIOR APPLICATION NUMBER: 60/259,434  
; PRIOR FILING DATE: 2000-12-27  
; PRIOR APPLICATION NUMBER: 09/948,993  
; PRIOR FILING DATE: 2001-09-06  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06  
; NUMBER OF SEQ ID NOS: 68  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 47  
; LENGTH: 12733  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Vector pPEPF1  
US-10-032-393-47

Query Match 5.2%; Score 52.2; DB 9; Length 12733;  
Best Local Similarity 42.9%; Pred. No. 0.0057;  
Matches 111; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

```
QY 72 CCCCCCTGNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 131
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4743 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 4684
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 132 GCCCCCCCCCGCAGCAGCTGTAGCCGCCAGCTCAGCCAGCTCAGCCAGCTCAGCCAGCT 191
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4683 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 4624
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 192 GCACACCACTCCCGGTGACGCCACGCCACCATGGACACAGCCCTGCGCCCTCGGCCCGG 251
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4623 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 4564
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 252 CTTTCTGCTGCTTCTACGACCATGTGACCCGACAGCCCTGCGCCCTCGGCCCGGCT 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4563 CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC 4504
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

QY 312 CCGGGCAGTACTGGGAC 330
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 4503 CCCCCCACTTCAAGGAC 4485
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

## RESULT 5

US-10-032-393-8/c

; Sequence 8, Application US/10032393  
; Publication No. US20030027286A1

; GENERAL INFORMATION:

; APPLICANT: Haselbeck, Robert

; APPLICANT: Wall, Daniel

; APPLICANT: Gross, Molly

; TITLE OF INVENTION: BACTERIAL PROMOTERS AND METHODS OF USE

; FILE REFERENCE: ELITRA.010A

; CURRENT APPLICATION NUMBER: US/10/032,393

; CURRENT FILING DATE: 2001-12-21

; PRIOR APPLICATION NUMBER: 60/259,434

; PRIOR FILING DATE: 2000-12-27

; PRIOR APPLICATION NUMBER: 09/948,993

; PRIOR FILING DATE: 2001-09-06

; PRIOR APPLICATION NUMBER: 60/230,335

; PRIOR FILING DATE: 2000-09-06

; NUMBER OF SEQ ID NOS: 68

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8

; LENGTH: 12739

; TYPE: DNA

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Vector pPEPF1

US-10-032-393-8

Query Match 5.2%; Score 52.2; DB 9; Length 12739;  
Best Local Similarity 42.9%; Pred. No. 0.0057;

Matches	111;	Conservative	0;	Mismatches	148;	Indels	0;	Gaps	0;
QY	72	CCCCCGCTNNNGG	131						
Db	4749	CCC							
QY	132	GGCCCCCGGCAGCAGTGTATGCCACGCTTCAGCCCGAGCTTCCAACAGAAGCAGCGCG	191						
Db	4689	CCC							
QY	192	GCACCACACTCCCGGTGACGGCCAGCCCATGTGACACAGACGCGCTTCGCGCTCCGCGCGG	251						
Db	4629	CCC							
QY	252	CTTTTCTCTGCTTTTATTACGACCATGTGACCCCGCACAGCGCTTCGCCCTGCGCTCGCCT	311						
Db	4569	CCC							
QY	312	CCCGGGCAGTACTGGGGAC	330						
Db	4509	CCCCCCCAC TTCAAAGGAC	4491						
<b>RESULT 6</b>									
US-10-095-407-16									
; Sequence 16, Application US/10095407									
; Patent No. US20020164330A1									
; GENERAL INFORMATION:									
; APPLICANT: Pan, Yang									
; TITLE OF INVENTION: NOVEL MOLECULES OF TANGO-77 RELATED PROTEIN FAMILY									
; FILE REFERENCE: 09404/052001									
; CURRENT APPLICATION NUMBER: US/10/095,407									
; PRIOR FILING DATE: 2002-03-11									
; PRIOR APPLICATION NUMBER: US 60/091,650									
; PRIOR FILING DATE: 1998-07-02									
; PRIOR APPLICATION NUMBER: US 60/054,646									
; NUMBER OF SEQ ID NOS: 18									
; SOFTWARE: FastSeq for Windows Version 3.0									
; SEQ ID NO 16									
; LENGTH: 152331									
; TYPE: DNA									
; ORGANISM: Homo sapiens									
; FEATURE:									
; NAME/KEY: misc_feature									
; LOCATION: (1)...(152331)									
; OTHER INFORMATION: n = A,T,C or G									
US-10-095-407-16									
Query Match 4.9%; Score 48.8; DB 9; Length 152331;									
Best Local Similarity 52.5%; Pred. No. 0.15;									
Matches	107;	Conservative	0;	Mismatches	97;	Indels	0;	Gaps	0;
QY	133	CCCCCGCGCAGCAGCTGTAGCCCGCAGCTTCAGCCCGAGCTTCCAACAGAAGCAGCGCGG	192						
Db	22019	CCCCCG							
QY	193	CACCCACATCCCGTGACGCGCGCCAGCGCAGATGTGACACAGCGCTTCGCGCTCCGCGCGCGG	252						
Db	22079	CCCCACCGGAGCG							
QY	253	TTTTTCTCTGCTTTTCTTACCGACCATGTGACCCCGCACAGCGCTTCGCGCTCCGCGCTCGC	312						
Db	22139	CGGCCCGGACGCCCGGACGCCCG							
QY	313	CCGGGCACTGTGGGACCTTCC	336						
Db	22199	CCCCCGCGGCG	22222						

RESULT 7  
US-10-078-090-83

```
; Sequence 83, Application US/10078090
; Publication No. US20030044815A1
; GENERAL INFORMATION:
; APPLICANT: Salceda, Susana
; APPLICANT: Macina, Roberto
; APPLICANT: Hu, Ping
; APPLICANT: Recipon, Hervé
; APPLICANT: Karra, Kalpana
; APPLICANT: Cafferkey, Robert
; APPLICANT: Sun, Yongming
; APPLICANT: Liu, Chenghua
; TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific Gene
; FILE REFERENCE: DEX-0312
; CURRENT APPLICATION NUMBER: US/10/078,090
; CURRENT FILING DATE: 2002-02-14
; PRIOR APPLICATION NUMBER: 60/268,999
; PRIOR FILING DATE: 2001-02-15
; NUMBER OF SEQ ID NOS: 210
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 83
; LENGTH: 954
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-078-090-83

Query Match      4.8%; Score 48.2; DB 9; Length 954;
Best Local Similarity 55.8%; Pred. No. 0.015;
Matches          92; Conservative    0; Mismatches   73; Indels     0; Gaps     0;

QY 701 TATTAAATCCGCTGGTTTGTGTTTTTAATTTTGTGTTTTTGAATTTTAAATAAG 760
Db 70  TTTTGTGTTTTTGTGTTTTTGTGTTTTTGTGTTTTTGTGTTTTTGTGTTTTTGTG 129

QY 761 AATTTCATTTAAGCACATTTACTGAAGCAATTGCTGCTGTATTTGGGGGAGCTG 820
Db 130 TTTTGTGTTTTTGTGTTTTTGTGTTTTTGTGTTTTTGTGTTTTTGTGTTTTTGTG 189

QY 821 GATCCACAGCTCGAGGGGTGGTGCCGGGGGAGGAGTGCTCGG 865
Db 190 GCCCGCGGGGGGGGGGGGGCCCCGGGGGGGGGGGGGGGG 234

RESULT 8
US-10-012-542-44/c
; Sequence 44, Application US/10012542
; Publication No. US20030044851A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 94 Human Secreted Proteins
; FILE REFERENCE: P2029P1
; CURRENT APPLICATION NUMBER: US/10/012,542
; CURRENT FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/461,325
; PRIOR FILING DATE: EARLIER FILING DATE: 1999-12-14
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,507
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,508
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,509
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/089,510
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-16
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,112
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-06-22
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/090,113
; NUMBER OF SEQ ID NOS: 532
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 44
; LENGTH: 569
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-012-542-44
```









GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:06:43 ; Search time 1290.57 Seconds  
(without alignments)  
12561.622 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_15000\_16000

Perfect score: 1001

Sequence: 1 gcctaaataatcccggtt.....ccctccctccactggagaa 1001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estha:\*

2: em\_esthum:\*

3: em\_estin:\*

4: em\_estmu:\*

5: em\_estov:\*

6: em\_estpl:\*

7: em\_estro:\*

8: em\_htc:\*

9: gb\_est1:\*

10: gb\_est2:\*

11: gb\_hic:\*

12: gb\_est3:\*

13: gb\_est4:\*

14: gb\_est5:\*

15: em\_estfun:\*

16: em\_estom:\*

17: gb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_Other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	715.8	71.5	1067	13	BM423874
2	692.6	69.2	1091	14	BM927280
3	676.4	67.6	1160	13	BM554172
4	651	65.0	654	14	BM722189
5	643.4	64.3	703	9	AI806984
6	639.6	63.9	654	14	BQ044936

7	628.6	62.8	954	14	BQ441751
8	602.6	60.2	978	9	AL522712
9	599	59.8	967	9	AL533351
10	597.8	59.7	931	9	AL531149
11	595.2	59.5	993	9	AL551103
12	593.8	59.3	626	9	AI963441
13	590.8	59.0	692	9	AA812217
14	589.2	58.9	1051	9	AL578663
15	572.8	57.2	602	10	AW118835
16	559.2	55.9	596	9	AI962618
17	550	54.9	585	13	BI770450
18	549.4	54.9	1164	12	BG028346
19	546.2	54.6	552	9	AI380422
20	542.8	54.2	673	10	BE383139
21	503.6	50.3	849	9	AL521191
22	500.8	50.0	583	9	AI401587
23	479	47.9	483	9	AI640571
24	451.8	45.1	563	10	AW130149
25	450.6	45.0	510	9	AI863571
26	443.4	44.3	509	9	AA811759
27	437.4	43.7	469	9	AA670427
28	432.2	43.2	450	10	AW118948
29	429.6	42.9	466	9	AA595224
30	417.4	41.7	505	9	AI824404
31	416.2	41.6	777	12	BG818326
32	407.6	40.7	425	13	BM148582
33	393	39.3	395	9	AI005074
34	391.6	39.1	792	12	BF347137
35	388	38.8	388	9	AI589149
36	380.8	38.0	404	14	BM717189
37	368	36.8	370	9	AI218536
38	347.6	34.7	542	10	AW104216
39	346.8	34.6	350	12	BF593529
40	336.6	33.6	352	12	BG389628
41	334	33.4	348	10	AW298630
42	324.2	32.4	393	10	AW236634
43	321.2	32.1	340	10	AW294932
44	318.8	31.8	515	10	AV609964
45	314.6	31.4	372	10	AW972482

## ALIGNMENTS

RESULT 1  
BM423874  
LOCUS BM423874 1067 bp mRNA linear EST 29-JAN-2002  
DEFINITION AGENCOURT\_6399121 NIH\_MGC\_41 Homo sapiens cDNA clone IMAGE:5517105  
5', mRNA sequence.  
ACCESSION BM423874  
VERSION BM423874.1 GI:18392086  
KEYWORDS EST  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 1067)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: DCTD/DP  
CDNA Library Preparation: Rubin Laboratory  
CDNA Sequencing by: The I.M.A.G.E. Consortium (LLNL)  
Clone Distribution: Agencourt Bioscience Corporation  
found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
Plate: L10M2018 row: m column: 10  
High quality sequence stop: 597.  
Location/Qualifiers 1. .1067

FEATURES  
source

```

:
:
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5517105"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/lab_host="DH10B (phage-resistant)"
/notes="CDNA: skin; Vector: pORF7; Site_1: XhoI; Site_2:
EcoRI; CDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
189 a 288 q 236 t
BASE COUNT

```

	Query Match	71.5%;	Score 715.8;	DB 13;	Length 1067;	
	Best Local Similarity	99.7%;	Pred. No. 8e-112;			
	Matches 717;	Conservative	0;	Mismatches	2;	Indels
			0;		0;	Gaps
QY	283	CCCGACACAGCCCTGCCCCACCTGCCCTCCGGGCGAGTACTGGGGAGACTTCCCTGGGGG	342			
Db	21					
		CCCTACACAGCCCTGCCCCACCTGCCCTCCGGGCGAGTACTGGGGAGACTTCCCTGGGGG	80			
QY	343	ACGGGAGGGAGGAGCAGCAGCTCTTTGGACAGAGGCTGGGCCCTCAGTGGACTGCCT	402			
Db	81					
		ACGGGAGGGAGGAGCAGCAGCTCTTTGGACAGAGGCTGGGCCCTCAGTGGACTGCCT	140			
QY	403	GCTCCACAGCCTGGGCTGACGTCCAGAGGCCGAGGCCAGGAACTCAGTGGAGGCCCTTGGT	462			
Db	141					
		GCTCCACAGCCTGGGCTGACGTCCAGAGGCCGAGGCCAGGAACTCAGTGGAGGCCCTTGGT	200			
QY	463	CCTGGGTCTCAGGATGGGTCTCTGGGGGCTCGTGTTCATCAAGACACGCCCTTGCACCAGC	522			
Db	201					
		CCTGGGTCTCAGGATGGGTCTCTGGGGGCTCGTGTTCATCAAGACACGCCCTTGCACCAGC	260			
QY	523	TCACCACATCTTCATCACCAGCAACGCGCAGGACTTGGCTCCGCCATCTCCAGAACTCAC	582			
Db	261					
		TCACCACATCTTCATCACCAGCAACGCGCAGGACTTGGCTCCGCCATCTCCAGAACTCAC	320			
QY	583	AAGCCATTGCTCCCCAGCTGGGGAACCTCAACCTCCCCCTGCCTCGGTTGGTGACAGAG	642			
Db	321					
		AAGCCATTGCTCCCCAGCTGGGGAACCTCAACCTCCCCCTGCCTCGGTTGGTGACAGAG	380			
QY	643	GGGTGGGACAGGGCGGGGGTTCCCTCTACATACCCTGCCATACCAACCCACGGTA	702			
Db	381					
		GGGTGGGACAGGGCGGGGGTTCCCTCTACATACCCTGCCATACCAACCCACGGTA	440			
QY	703	TTAATTCGCTGTTTTGTTTTTATTTAATTTTTTTGTTTTGATTTTTTTAATAAGAA	762			
Db	441					
		TTAATTCGCTGTTTTGTTTTTATTTAATTTTTTTGTTTTGATTTTTTTAATAAGAA	500			
QY	763	TTTTCAATTTAAGACATTTATCTGAAGAAATTTGCTGTGTATTTGGGGGAGCTTGA	822			
Db	501					
		TTTTCAATTTAAGACATTTATCTGAAGAAATTTGCTGTGTATTTGGGGGAGCTTGA	560			
QY	823	TCAGAGCTGGAGGGTGGTCCGGGGAGGAGTGGCTCGGAAGGGGCCCCACTCTC	882			
Db	561					
		TCAGAGCTGGAGGGTGGTCCGGGGAGGAGTGGCTCGGAAGGGGCCCCACTCTC	620			
QY	883	CTTTCAGTCCCTGTGCCCCCCAGTCTCCTCCTCAGCCTTTTCCTCCTCAGTTTCTCT	942			
Db	621					
		CTTTCAGTCCCTGTGCCCCCCAGTCTCCTCCTCAGCCTTTTCCTCCTCAGTTTCTCT	680			
QY	943	TTAAAACTGTGAAGTACTACTTTCCAGAGCCTTGCCTTCCCTCCCTCCACATGGAGAA	1001			
Db	681					
		TTAAAACTGTGAAGTACTACTTTCCAGAGCCTTGCCTTCCCTCCCTCCACATGGAGAA	739			

RESULT 2			
BM927280			
LOCUS	BM927280	1091 bp	linear
		mrna	EST 12-MAR-2002

DEFINITION	AGENCOURT_6688399 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5768262 5', mRNA sequence.
ACCESSION	BM927280
VERSION	BM927280.1 GI:19377659
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS	1 (bases 1 to 1091)
TITLE	NIH-MGC <a href="http://mgc.nci.nih.gov/">http://mgc.nci.nih.gov/</a> .
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)
COMMENT	Contact: Robert Strausberg, Ph.D. Email: <a href="mailto:cgapbs-remail.nih.gov">cgapbs-remail.nih.gov</a> Tissue Procurement: Life Technologies, Inc. cDNA Library Preparation: Life Technologies, Inc. cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Agencourt Bioscience Corporation Clone distribution by: Agencourt Bioscience Corporation found through the I.M.A.G.E. Consortium <a href="http://image.llnl.gov">http://image.llnl.gov</a> Plate: LLNL12827 row: n column: 08 High quality sequence stop: 621.

<b>FEATURES</b>	<b>SOURCE</b>
-----------------	---------------

FEATURES	source
Location/Qualifiers	
1..1091	
/organism="Homo sapiens"	
/db_xref="taxon:9606"	
/clone="IMAGE:5768263"	
/clone_lib="NIH_MGC_121"	
/lab_host="DH10B"	
/notes="Organ: brain; Vector: pCMV-SPORT6; Site.1': NotI; Site.2': EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 Kb, insert size range 0.7-3.5 Kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH_MGC Library."	
278 a	339 c 286 q 188 t
BASE COUNT	

BASE COUNT	278 a	339 c	286 g	188 t	thus is a NIM_MGC Library.
ORIGIN					
Query Match	69.2%	Score 692.6;	DB 14;	Length 1091;	
Best Local Similarity	97.8%;	Pred. No. 6.7e-108;			
Matches 71;	Conservative	0;	Mismatches 14;	Indels 2;	Gaps 1;
QY	131	GGCCCCCGCCGACGAGCTGTAGCCCGCAGCTCAGCGCCCGACAGCTCCAAACAGAACGACGCC	190		
Db	3	GGCCCCCGCCGACGAGCTGTAGCCCGCAGCTCAGCGCCCGACAGCTCCAAACAGAACGACGCC	62		
QY	191	GGCCACCGACTCCCGCTGACCGCCACGCGCCACATGGACACAGCCCTCGCCCTCGGCCCGG	250		
Db	63	GGCCACCGACTCCCGCTGACCGCCACGCGCCACATGGACACAGCCCTCGCCCTCGGCCCGG	122		
QY	251	GCCTTTCTCTGCTTTCTACCGACCATGTGACCCCGCACACAGCCCTGCCGCCACCTGC	310		
Db	123	GCCTTTCTCTGCTTTCTACCGACCATGTGACCCCGCACACAGCCCTGCCGCCACCTGC	182		
QY	311	TCCCGGCGAGTACTGGGGACCTTCCCTGGGGGACGGGGAGGGAGGACGAGCTCCTT	370		
Db	183	TCCCGGCGAGTACTGGGGACCTTCCCTGGGGGACGGGGAGGGAGGAGGACTCCTT	242		
QY	371	GGACAGAGGCTTGGCCCTCAGTGGAGTGCTGTCTCCACAGCCTGGGCTCAGGTCAG	430		
Db	243	GGACAGAGGCTTGGCCCTCAGTGGAGTGCTGTCTCCACAGCCTGGGCTCAGGTCAG	302		
QY	431	GCCGAGGCCAGGAACGTAGTGAAGGCCCTCGTCTGGTCTCAGGATGGGTCTCGGGGC	490		
Db	303	GCCGAGGCCAGGAACGTAGTGAAGGCCCTCGTCTGGTCTCAGGATGGGTCTCGGGGC	362		
QY	491	CTCGTGTTATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAACGC	550		

```

Db 363 CTCGCTGTCATCAAGACACCCCTCTGCCAGCTCACACACTCTTCATCACCAGCAACGC 422
QY 551 CAGGACTTGGCTCCCTCCCTCCTCAGAACTCACAAGCCATGCTCCCTCAGCTGGGAACCT 610
Db 423 CAGGACTTGGCTCCCTCCCTCCTCAGAACTCACAAGCCATGCTCCCTCAGCTGGGAACCT 482
QY 611 CAACTTCCCTCCCTGCTCGCTGGTGGTGACAGAGGGGTGGGACAGGGGGGGTTCCTCC 670
Db 483 CAACTTCCCTCCCTGCTCGCTGGTGGTGACAGAGGGGTGGGACAGGGGGGGTTCCTCC 542
QY 671 CTGTACATACCTTGCATACCAACCCAGGTATTAATTCCTCGCTGGTGGTGGTGGTGGT 730
Db 543 CTGTACATACCTTGCATACCAACCCAGGTATTAATTCCTCGCTGGTGGTGGTGGTGGT 602
QY 731 TAAATTTTCTTTGATTTTATTAAGAAATTTTCAATTTAAGCACAATTTATATCTGAA 790
Db 603 TAAATTTTCTTTGATTTTATTAAGAAATTTTCAATTTAAGCACAATTTATATCTGAA 662
QY 791 GGAATTTCTGTGATTTGGGGG--AGCTGGATCCAGAGCTGGAGGGGGTGGTTCGG 848
Db 563 GGAATTTCTGTGATTTGGGGGAACTGGAATCCAAAGCTGGAGGGGGGGGACCCC 722
QY 849 GGGAGGAG 857
Db 723 GAGCTAGAG 731

RESULT 3
BM554172
LOCUS
DEFINITION
AGENCOURT_5581002 NIH_MGC_41 Homo sapiens cdna clone IMAGE:5469285
5', mRNA sequence.
ACCESSION
BM554172
VERSION
BM554172.1 GI:18793541
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NIH-MGC http://img.nci.nih.gov/,
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: DCTP/DTP
cdna Library Preparation: Rubin Laboratory
cdna Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1CM1974 row: d column: 22
High quality sequence stop: 593.
FEATURES
Location/Qualifiers
1..1160
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5469285"
/clone_lib="NIH_MGC_41"
/tissue_type="amelanotic melanoma, cell line"
/note="Organ: skin; Vector: pOTB7; Site:1: XhoI; Site:2:
EcoRI; cdna made by oligo-dr priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGAG(G). Library constructed by Ling Hong in the
laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cdna synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT
173 a 324 c 280 g 219 t 164 others
ORIGIN

```

```

Query Match 67.6%; Score 676.4; DB 13; Length 1160;
Best Local Similarity 97.8%; Pred. No. 3.7e-105;
Matches 707; Conservative 0; Mismatches 12; Indels 4; Gaps 2;

QY 283 CCCGACACAGCCCTGCCCCACCTGCCCCCTCCGGGACAGTCTCCCTGGGGG 342
Db 13 CTCGACACAGCCCTGCCCCACCTGCCCCCTCCGGGACAGTCTCCCTGGGG 72
QY 343 ACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 402
Db 73 ACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 132
QY 403 GCTCCACACGCCCTGGGCTGAGCTCAGAGGCGGAGGAGGAGGAGGAGGAGG 462
Db 133 GCTCCACACGCCCTGGGCTGAGCTCAGAGGCGGAGGAGGAGGAGGAGGAGG 192
QY 463 CTTGGGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 522
Db 193 CTTGGGCTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 252
QY 523 TCACCAATCTTTCATCACCAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 582
Db 253 TCACCAATCTTTCATCACCAGCAAGAGGAGGAGGAGGAGGAGGAGGAGGAGG 312
QY 583 AAGCCATGCTCCCCAGCTGGGGAACCTCAACCTCCCCCTGCTCGGTTGGTGACAG 642
Db 313 AAGCCATGCTCCCCAGCTGGGGAACCTCAACCTCCCCCTGCTCGGTTGGTGACAG 372
QY 643 GGGGTGGGACAGG-GGCGGGGGTTCCTCCCTGTACATACCTCCCATACCAACCCAGGT 701
Db 373 GGGGTGGGACAGGNGGGGGGGTTCCTCCCTGTACATACCTCCCATACCAACCCAGGT 432
QY 702 ATTAATCTCTGCTGTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 761
Db 433 ATTAATCTCTGCTGTTTGTATTTTATTTTATTTTATTTTATTTTATTTTATTTT 492
QY 762 ATTTTCATTTAAGCACAATTTATCTAGAGGAATTTGCTGTGTATTTGGGGGAGCTGG 821
Db 493 ATTTTCATTTAAGCACAATTTATCTAGAGGAATTTGCTGTGTATTTGGGGGAGCTGG 552
QY 822 ATCCAGAGCTGGAGGGGGTGGTCCGGG---GAGGGAGTGTCTCGAAGAGGGGCCCCAC 878
Db 553 ATCCAGAGCTGGAGGGGGTGGTCCGGGAGGAGTGGGNTCGAAGAGGGGCCCCAC 612
QY 879 TCTCCTTTCATGTCCTGTCGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 938
Db 613 TCTCCTTTCATGTCCTGTCGCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCTCCCT 672
QY 939 CTCTTTAAACTGTGAAGTACTAACTTTCCAAAGGCTGCTTCCCTCCCTCCCTCCCTCCCT 998
Db 673 CTCTTTAAACTGTGAAGTACTAACTTTCCAAAGGCTGCTTCCCTCCCTCCCTCCCTCCCT 732
QY 999 GAA 1001
Db 733 GAA 735

RESULT 4
BM722189
LOCUS
DEFINITION
BM722189 UI-E-E00-ahy-m-11-0-UI.r1 UI-E-E00 Homo sapiens cdna clone
654 bp mRNA linear EST 01-MAR-2002
ACCESSION
BM722189 UI-E-E00-ahy-m-11-0-UI 5', mRNA sequence.
VERSION
BM722189.1 GI:19042595
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 654)
AUTHORS
Ronald M.F., Lennon, G. and Soares, M.B.
TITLE
Normalization and subtraction: two approaches to facilitate gene

```



```
QY 131 GGGCCCCCGCCGAGGAGCTGTAGCCCCAGCCCTCAGCCCGCCAGCTCCCAACAGAGCAGCCCC 190
Db 655 GGGCCCCCGCCGAGGAGCTGTAGCCCCAGCCCTCAGCCCGCCAGCTCCCAACAGAGCAGCCCC 596
QY 191 GGGCCACCACCTCCCGTACCGCCCGCCAGCCACATGATGACACAGCCCTCGCCCTCCGCCCCG 250
Db 595 GGGCCACCACCTCCCGTACCGCCCGCCAGCCACATGATGACACAGCCCTCGCCCTCCGCCCCG 536
QY 251 GCTTTCTCTGCTTCTTACCGACCATGTGACCCCGCCAGCCAGCCCTGCCCCCAGCCCTGCC 310
Db 535 GCTTTCTCTGCTTCTTACCGACCATGTGACCCCGCCAGCCAGCCCTGCCCCCAGCCCTGCC 476
QY 311 TCCGGGAGTACTGGGAGCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 370
Db 475 TCCGGGAGTACTGGGAGCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 416
QY 371 GGACAGAGGCGCTGGGCGCTCAGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 430
Db 415 GGACAGAGGCGCTGGGCGCTCAGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 356
QY 431 GCCGAGGCGGAGAACTAGTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 490
Db 355 GCCGAGGCGGAGAACTAGTGAGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 296
QY 491 CTGCTGTTTCATCAGACACCCCTCTGCCCCAGCTCAACACATCTTTCATCACCAGCAAGCG 550
Db 295 CTGCTGTTTCATCAGACACCCCTCTGCCCCAGCTCAACACATCTTTCATCACCAGCAAGCG 236
QY 551 CAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 610
Db 235 CAGGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 176
QY 611 CAACCTCCCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 670
Db 175 CAACCTCCCGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 116
QY 671 CTGTATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 730
Db 115 CTGTATACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 56
QY 731 TAATTTTTTTGTTGTTGTTTATTAAGAATTTTTCATTTTAAAGCATTTTAA 785
Db 55 TAATTTTTTTGTTGTTGTTTATTAAGAATTTTTCATTTTAAAGCATTTTAA 1

RESULT 6
BQ044936/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

654 bp mRNA linear EST 28-MAR-2002
UI-H-EU0-azo-e-05-0-UI.s1 NCI-CGAP_Carl Homo sapiens cdna clone
IMAGE:5851372 3', mRNA sequence.
BQ044936
EST.
human.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 654)
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished (1997)
Contact: Robert Strausberg, Ph.D.
Email: cgapb-r@mail.nih.gov
Tissue Procurement: Dr. Jose Mercuende
cdna Library preparation: Dr. M. Bento Soares, University of Iowa
cdna Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
The following repetitive elements were found in this cdna
sequence: 1-73, >AT rich#Low_complexity (matched complement)
Seq primer: M13 FORWARD
```

FEATURES  
Source  
Location/Qualifiers  
1..654

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5851372"  
/tissue\_type="Osteoarthritic Cartilage"  
/dev\_stage="Adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Organ: Knee; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; NCI-CGAP\_Carl is a cdna library containing the following tissue(s): Osteoarthritic Cartilage The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cdna synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cdna was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cdna contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is TGTATCAGCTT.

BASE COUNT 135 a 161 c 227 g 131 t  
ORIGIN  
TAG\_LTB=UI-H-EU0  
TAG\_TISSUE=Osteoarthritic cartilage  
TAG\_SEQ=TGATCAGCT

Query Match 63.9%; Score 639.6; DB 14; Length 654;  
Best Local Similarity 98.6%; Pred. No. 7, 1e-99;  
Matches 645; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 141 CCAGGACGCTGTAGCCCCAGCTCAGCCCCAGCTCAACAGAGCAGCGGCCACCCAC 200  
Db 654 CCAGGACGCTGTAGCCCCAGCTCAGCCCCAGCTCAACAGAGCAGCGGCCACCCAC 595  
QY 201 TCCCGCTGACCGCCCGCCAGCCATGACACAGCCCTCGCCCTCCGCCCGGCTTTCTCT 260  
Db 594 TCCCGCTGACCGCCCGCCAGCCATGACACAGCCCTCGCCCTCCGCCCGGCTTTCTCT 535  
QY 261 GCCTTTTACGACCATGTGACCCCGCCAGCCCTGCCCCAGCTGCCCTCCCGGGCAG 320  
Db 534 GCCTTTTACGACCATGTGACCCCGCCAGCCCTGCCCCAGCTGCCCTCCCGGGCAG 475  
QY 321 TACTGGGAGCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 380  
Db 474 TACTGGGAGCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 415  
QY 381 CTGGGCGCTCAGTGGAGCTGCTGCTGCCACAGCCCTGGGCTGACGTCAGAGGCCGAGGCCA 440  
Db 414 CTGGGCGCTCAGTGGAGCTGCTGCTGCCACAGCCCTGGGCTGACGTCAGAGGCCGAGGCCA 355  
QY 441 GGAAGTGAAGTGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 500  
Db 354 GGAAGTGAAGTGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 295  
QY 501 TCAAGACACCCCTCTGCCAGCTCACCACATCTTTCATCACCAGCAAGCGGCCAGACTGG 560  
Db 294 TCAAGACACCCCTCTGCCAGCTCACCACATCTTTCATCACCAGCAAGCGGCCAGACTGG 235  
QY 561 CTCGCCCATCTCAGAACTCACAAGCCATTTGCTCCCGAGTGGGGAACCTCAACCTCCGC 620  
Db 234 CTCGCCCATCTCAGAACTCACAAGCCATTTGCTCCCGAGTGGGGAACCTCAACCTCCGC 175  
QY 621 CTGCTCCTCGTGTGGTGACAGAGGGGTGGACAGGGCGGGGGTTCCTCCCTGTACATAC 680  
Db 174 CTGCTCCTCGTGTGGTGACAGAGGGGTGGACAGGGCGGGGGTTCCTCCCTGTACATAC 115  
QY 681 CTGCTCCTACCAACCCAGGATTAATTTCTCGCTGGTTTGTGTTTATTTATTTT 740  
Db 114 CTGCTCCTACCAACCCAGGATTAATTTCTCGCTGGTTTGTGTTTATTTATTTT 55

```

QY 741 GTTTTGAATTTTAAATAAGATTTTCATTTAAGACACATTTATATACTGAAGAA 794
|||||
Db 54 GTTTTGAATTTTAAATAAGATTTTCATTTAAGACACAAAAAATAAAAAA 1
|||||

RESULT 7
BQ441751 954 bp mRNA linear EST 24-MAY-2002
LOCUS AGENCOURT_7761844 NIH_MGC_70 Homo sapiens cDNA clone IMAGE:6018014
DEFINITION 5'; mRNA sequence.
ACCESSION BQ441751
VERSION BQ441751.1 GI:21180827
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 954)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: rcgaps@mail.nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L1AM13218 row: d column: 15
High quality sequence stop: 563.
FEATURES
Location/Qualifiers
1..954
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6018014"
/clone_lib="NIH_MGC_70"
/tissue_type="epithelioid carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: pancreas; Vector: pCMV-SPORT6; Site:1: NotI;
Site:2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.1 kb. Library constructed by Life
Technologies."
BASE COUNT 149 a 298 C 271 g 235 t 1 others
ORIGIN
Query Match 62.8%; Score 628.6; DB 14; Length 954;
Best Local Similarity 98.4%; Pred. No. 4.8e-97;
Matches 634; Conservative 0; Mismatches 10; Indels 0; Gaps 0;

QY 309 CCTCCGGGCGAGTACTGGGACCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGACTCC 368
|||||
Db 1 CCTCCGGGCGAGTACTGGGACCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGACTCC 60
|||||

QY 369 TTGACAGAGGCTGGGCGCTCAGTGACCTGCCTGCCACAGCCCTGGGCTGACGTCAG 428
|||||
Db 61 TTGACAGAGGCTGGGCGCTCAGTGACCTGCCTGCCACAGCCCTGGGCTGACGTCAG 120
|||||

QY 429 AGGCCAGGCCAGGAACATGATGAGGCGCCCTGCTGGTCTCAGGATGGGCTCGGCGG 488
|||||
Db 121 AGGCCAGGCCAGGAACATGATGAGGCGCCCTGCTGGTCTCAGGATGGGCTCGGCGG 180
|||||

QY 489 GCTCTGTTTCATCAGACACACCCCTCTGCCAGCTACACATCTTCATCACCAGCAAC 548
|||||
Db 181 GCTCTGTTTCATCAGACACACCCCTCTGCCAGCTACACATCTTCATCACCAGCAAC 240
|||||

QY 549 GCAGGACTTGGTCTCCCACTCTCAGAACTCAGAACTTGTCTCCCACTGGGGAAC 608
|||||
Db 241 GCAGGACTTGGTCTCCCACTCTCAGAACTCAGAACTTGTCTCCCACTGGGGAAC 300
|||||

QY 609 CTCACCTCCCTCGCTCGGTTGGTGACAGAGGGGTTGGGACAGGGGGGGGGTTC 668
|||||

```

```

Db 301 CTCACCTCCCTCGCTCGGTTGGTGACAGAGGGGTTGGGACAGGGGGGGGTCC 360
|||||
QY 669 CCTGTACATACCTCGCATACCAACCCAGGTATTAATCTCGCTGTTTGTGTTTAT 728
|||||
Db 361 CCTGTACATACCTCGCATACCAACCCAGGTATTAATCTCGCTGTTTGTGTTTAT 420
|||||
QY 729 TTAAATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 788
|||||
Db 421 TTAAATTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTG 480
|||||
QY 789 AAGGAATTTGTCTGTATTTGGGGGAGCTGATCCAGAGCTGGAGGGGGTGGTCCGG 848
|||||
Db 481 AAGGAATTTGTCTGTATTTGGGGGAGCTGATCCAGAGCTGGAGGGGGTGGTCCGG 540
|||||
QY 849 GGGAGGAGTGGTCTGGAAGGGGCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 908
|||||
Db 541 GGGAGGAGTGGTCTGGAAGGGGCCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 600
|||||
QY 909 CTCCTCTCTCAGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 952
|||||
Db 601 CGGCTCTCAGCTTTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 644
|||||

RESULT 8
AL522712 978 bp mRNA linear EST 13-FEB-2001
LOCUS AL522712 LTI_NFL004_NBC2 Homo sapiens cDNA clone CS0DB009G04 3
DEFINITION prime, mRNA sequence.
ACCESSION AL522712
VERSION AL522712.1 GI:12786205
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 978)
AUTHORS L.I.W.B., Gruber,C., Jessee,J. and Polayes,D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
Bp 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
Location/Qualifiers
1..978
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DB009G04"
/clone_lib="LTI_NFL004_NBC2"
/sex="male"
/tissue_type="neuroblastoma cells"
/lab_host="DH10B"
/notes="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed
by Life Technologies. Contact : Feng Liang Life
Technologies, a division of Invitrogen 9800 Medical Center
Drive Rockville, Maryland 20850, USA Fax : (1) 301 610
8371 Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 167 a 293 c 331 g 177 t 10 others
ORIGIN
Query Match 60.28; Score 602.6; DB 9; Length 978;
Best Local Similarity 90.9%; Pred. No. 1.2e-92;
Matches 613; Conservative 5; Mismatches 55; Indels 1; Gaps 1;

QY 77 CGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 136
|||||
Db 673 CCTGGACACTGTAGCGGACAGCGGGGGTGGGGGCGGAGCGGGGTGGCCTGGCCCC 614
|||||

```



```

QY 137 CCGCCAGGAGCTGTAGCCCGAGCCTCAGCCCGAGCTCCACAGAGCAGCCGCGCCAC 196
Db 613 CCGCCAGGAGCTGTAGCCCGAGCCTCAGCCCGAGCTCCACAGAGCAGCCGCGCCAC 554
QY 197 CCACCTCCCGTACCGCCCGCCAGCCACATGACACAGCCCTCGCCCTCCGCGCGGCTTTT 256
Db 553 CCACCTCCCGTACCGCCCGCCAGCCACATGACACAGCCCTCGCCCTCCGCGCGGCTTTT 494
QY 257 CTCCTGCCCTTACCGACCATGTGACCCCGCAGCCCTGCCCCACCTGCCCCCANCY-CCCTCCCG 316
Db 493 CTCCTGCCCTTACCGACCATGTGACCCCGCAGCCCTGCCCCCANCY-CCCTCCCG 435
QY 317 GCAGTACTGGGACCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376
Db 434 GCAGTACTGGGACCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 375
QY 377 AGGCTTGGGCGCTCAGTGAGCTGCTGCTCCACAGCCCTGCCCCACCTGCCCCCANCY-CCCTCCCG 436
Db 374 AGGCTTGGGCGCTCAGTGAGCTGCTGCTCCACAGCCCTGCCCCACCTGCCCCCANCY-CCCTCCCG 315
QY 437 GCCAGGAATCAGTGAGGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 496
Db 314 GCCAGGAATCAGTGAGGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 255
QY 497 TCCATCAGACACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 556
Db 254 TCCATCAGACACCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 195
QY 557 TTGGCTCCCGCTCAGTACTCAGAGCTCAGAGCTCAGAGCTCAGAGCTCAGAGCTCAGAGCTC 616
Db 194 TTGGCTCCCGCTCAGTACTCAGAGCTCAGAGCTCAGAGCTCAGAGCTCAGAGCTCAGAGCT 135
QY 617 CCCCCCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 676
Db 134 CCCCCCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 75
QY 677 ATACCTGCTCAGTACTCAGAGCTCAGAGCTCAGAGCTCAGAGCTCAGAGCTCAGAGCTC 736
Db 74 ATACCTGCTCAGTACTCAGAGCTCAGAGCTCAGAGCTCAGAGCTCAGAGCTCAGAGCTC 736
QY 737 TTTTGTGTTTGAATTT 750
Db 14 TTTTGTGTTTGAATTT 1

```

```

RESULT 9
AL533351
LOCUS
DEFINITION
AL533351 Lr1_FL015_Brn1 Homo sapiens cDNA clone CS0DN004Y122 5
prime, mRNA sequence.
ACCESSION
AL533351
VERSION
AL533351.1 GI:12796844
KEYWORDS
EST.
SOURCE
human.
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 967)
AUTHORS
Li.W.B., Gruber,C., Jesse,J. and Polayes,D.
TITLE
Full-length cDNA libraries and normalization
JOURNAL
Unpublished (2001)
COMMENT
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.
FEATURES
source
1. .967
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DN004Y122"
/clone_lib="Lr1_FL015_Brn1"
/sex="male"

```

```

/tissue_type="Adult brain"
/note="Vector: pcwvSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pcwvSPORT 6
vector. Library was constructed by Life Technologies.
Contact : Feng Liang Life Technologies, a division of
Invitrogen 9800 Medical Center Drive Rockville, Maryland
20850, USA Fax : (1) 301 610 8371 Email :
filiang@lifetech.com URL :
http://fulllength.invitrogen.com"
BASE COUNT 173 a 325 c 286 g 177 t 6 others
ORIGIN

```

```

Query Match 59.8%; Score 599; DB 9; Length 967;
Best Local Similarity 90.9%; Pred. No. 4.9e-92;
Matches 621; Conservative 5; Mismatches 55; Indels 2; Gaps 2;

```

```

QY 77 CGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 136
Db 285 CTTGGACACTCTGAGCGGACAGCGCGGGGGTGGGGGGGAGCGGGGTGGCTGGCTGGCCCC 344
QY 137 CCGCGCCAGGAGCTGTAGCCCGCAGCCCTCAGCCCGCAGCTCCACAGAGAGCAGCCGCGCCAC 196
Db 345 CCGCGCCAGGAGCTGTAGCCCGCAGCCCTCAGCCCGCAGCTCCACAGAGAGCAGCCGCGCCAC 404
QY 197 CCACCTCCCGCTGACCGCCCGCAGCGCACATGACAGCCCTCGCCCTCGCCCTCGCCCTCGCT 256
Db 405 CCACCTCCCGCTGACCGCCCGCAGCGCACATGACAGCCCTCGCCCTCGCCCTCGCCCTCGCT 464
QY 257 CTCCTGCTTTTACCGACCATGTGACCCCGCAGCGCCCTCGCCCTCGCCCTCGCCCTCGCT 316
Db 465 CTCCTGCTTTTACCGACCATGTGACCCCGCAGCGCCCTCGCCCTCGCCCTCGCCCTCGCT 524
QY 317 GCAGTACTGGGAGCCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 376
Db 525 GCAGTACTGGGAGCCTTCCTCGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 584
QY 377 AGGCTTGGGCGCTCAGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 436
Db 585 AGGCTTGGGCGCTCAGTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 644
QY 437 GCCAGGAATCAGTGAGGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 495
Db 645 GCCAGGAATCAGTGAGGCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 704
QY 496 GTTCATCAAGACACCCCTCTGCGCCAGCTCACCACATCTTCATCACCAGCAAGCCAGGA 555
Db 705 GTTCATCAAGACACCCCTCTGCGCCAGCTCACCACATCTTCATCACCAGCAAGCCAGGA 764
QY 556 CTTGGCTCCCGCTCCTCAGAACTCAGAGCCATTTGCTCCCGAGCTGGGGAACTCAACC 615
Db 765 CTTGGCTCCCGCTCCTCAGAACTCAGAGCCATTTGCTCCCGAGCTGGGGAACTCAACC 824
QY 616 TCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 674
Db 825 TCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 884
QY 675 ACATACCTGCCATCACCACCCAGGATTAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 734
Db 885 ACATACCTGCCATCACCACCCAGGATTAATTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 944
QY 735 TTTTGTGTTTGAATTTTAAAT 757
Db 945 TTTTGTGTTTGAATTTTAAAT 967

```

```

RESULT 10
AL531149/c
LOCUS
DEFINITION
AL531149 Lr1_NFL001_NBC4 Homo sapiens cDNA clone CS0DN001Y13 3
prime, mRNA sequence.
ACCESSION
AL531149

```



AL5311149.1 GI:12794642  
EST.  
SOURCE Homo sapiens  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 931)  
L4.W.B., Gruber,C., Jesse,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
Location/Qualifiers  
1..931  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DM001YN13"  
/clone\_lib="LTI\_NFL001\_NBC4"  
/sex="male"  
/tissue\_type="neuroblastoma cells"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cdna was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cdna was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 166 a 275 c 316 g 172 t 2 others

ORIGIN

Query Match 59.78; Score 597.8; DB 9; Length 931;  
Best Local Similarity 92.09; Pred. No. 7.8e-92;  
Matches 621; Conservative 0; Mismatches 52; Indels 2; Gaps 2;

QY 77 CGTNNNGGCGCC 136  
| | | | |  
DB 677 CTTGGACACTGTAGCGGCACAGCGGGGGTGGGGGGGGACGGGGGTGCCCTGGCGCC 618  
| | | | |  
QY 137 CCCGCCAGGACGTGTAGCCCCCAGCTCAGCCCCCAGCTCCAACAAGACGCCCGGCCAC 196  
| | | | |  
DB 617 CCCGCCAGGACGTGTAGCCCCCAGCTCAGCCCCCAGCTCCAACAAGACGCCCGGCCAC 558  
| | | | |  
QY 197 CCAGTCCTCC-CGTGACGCCCGCACCATGACACAGCCCTCGCCCTCCGCCCGCGCTTT 255  
| | | | |  
DB 557 CCATCCCGGGTGACCGCCCAACGCATGACACAGCCCTCGCCCTCCGCCCGCGCTTT 498  
| | | | |  
QY 256 TCTTGCTTTTCTACCAGCATGTGACCCCGACAGCCCTCGCCCGACCTGCCTCCCG 315  
| | | | |  
DB 497 TCCTGCCTTTCTACCGACCATGTGACCCCGACAGCCCTGCCCCCACCCTGCCCTCCCG 438  
| | | | |  
QY 316 GGCACTACTGGGGACCTTCCTTGGGGACGGGGAGGAGGAGGACGACCTCCTTTGGACA 375  
| | | | |  
DB 437 GGCACTACTGGGGACCTTCCTTGGGGACGGGGAGGAGGAGGACGACCTCCTTTGGACA 378  
| | | | |  
QY 376 GAGGCTTGGGCCCCTCAGTGACCTCCCTGCCACAGCCCTGGGCTGAGCTCAGAGCGGA 435  
| | | | |  
DB 377 GAGGCTTGGGCCCCTCAGTGACCTCCCTGCCACAGCCCTGGGCTGAGCTCAGAGCGGA 318  
| | | | |  
QY 436 GGCCAGGAACCTGAGTAGGGCCCTGTGCTGGGTCTCAGATGGGTCTTGGGGGCTCGT 495  
| | | | |  
DB 317 GGCCAGGAACCTGAGTAGGGCCCTGTGCTGGGTCTCAGATGGGTCTTGGGGGCTCGT 258  
| | | | |  
QY 496 GTTCATCAAGACACCCCTCTGCCCAGCTCAACACATCTTCATCAACAGC-AAAGCCGAGG 554  
| | | | |  
DB 257 GTTCATCAAGACACCCCTCTGCCAGCTCAACACATCTTCATCAACAGCTAAAGCCGAGG 198  
| | | | |  
QY 555 ACTTGGTCCCCATCCTCAGAATCAAGACCATTTGCTCCCCAGCTGGGGAACCTCAAC 614  
| | | | |



Email: cgapbs-r@mail.nih.gov  
Tissue Procurement: Louis M. Staudt, M.D., Ph.D., David Allman, Ph.D., Gerald Marti, M.D.  
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.  
CDNA Library Arrayed by: Greg Lennon, Ph.D.  
DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
[www-bio.llnl.gov/bbrp/image/image.html](http://www-bio.llnl.gov/bbrp/image/image.html)  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 469.

FEATURES	SOURCE
----------	--------

```

Location/Qualifiers
1. 1. 962
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:138100"
/tissue_type="NCI_CGAP_GCB1"
/lab_host="DH10B"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was prepared from human tonsillar cells enriched for
germinal center B cells by flow sorting (CD20+, IgD+),
provided by Dr. Louis M. Staudt (NCI), Dr. David Allman
(NCI) and Dr. Gerald Marti (CBER). cDNA synthesis was
primed with a Not I - oligo(dT) primer
[5'-GTGTACCAATCGTAGCGGCGCGCTCATTTTTTTTTTTT-3'.
]. Double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. Library
went through one round of normalization, and was
constructed by Bento Soares and M. Fatima Bonaldo."
143 a 183 c 232 g 131 t 3 others

```

Query Match	59.0%;	Score	590.8;	DB	9;	Length	692;	
Best Local Similarity	97.2%;	Pred. No.	1.3e-90;					
Matches	632;	Conservative	0;	Mismatches	14;	Indels	4;	
							Gaps	3;
<hr/>								
QY	131	GGCCCCCCCCG	CAGGAGCTGTAG	CCCCAGCCTCAG	CCCCAGCTCCAC	AAGACGACGCC	190	
Db								
QY	131	GGCCCCCCCCG	CAGGAGCTGTAG	CCCCAGCCTCAG	CCCCAGCTCCAC	AAGACGACGCC	190	
Db								
QY	191	GGCCACCCACT	CCCCGCTGAC	CGGCCA-CGGCCACAT	GGACACAGCCCTCG	CCCTCGCCCTCG	249	
Db								
QY	191	GGCCACCCACT	CCCCGCTGAC	CGGCCA-CGGCCACAT	GGACACAGCCCTCG	CCCTCGCCCTCG	249	
Db								
QY	250	GGCTTTTCT	CTGCTTTTCTAC	GACCATGTGAC	CCCGCACAGCCCTG	CCCCACACCTG	309	
Db								
QY	250	GGCTTTTCT	CTGCTTTTCTAC	GACCATGTGAC	CCCGCACAGCCCTG	CCCCACACCTG	309	
Db								
QY	310	CTCCCGGGCAG	TACTGGGGACCTTC	CCCTGGGGACCGGG	AGGAGGAGCGAGG	ACTCTCT	369	
Db								
QY	310	CTCCCGGGCAG	TACTGGGGACCTTC	CCCTGGGGACCGGG	AGGAGGAGCGAGG	ACTCTCT	369	
Db								
QY	370	TGGACAGAGGC	CTGGGCCCTCAGT	GGACTGCCTGCTC	CCACAGCCCTG	GGCTGACGTCAGA	429	
Db								
QY	370	TGGACAGAGGC	CTGGGCCCTCAGT	GGACTGCCTGCTC	CCACAGCCCTG	GGCTGACGTCAGA	429	
Db								
QY	430	GGCCGAGGGC	AGGAATGATGAG	GGCCCCCTGCTG	CTGGTCTCAGGATGGG	CTCTGGGGG	489	
Db								
QY	430	GGCCGAGGGC	AGGAATGATGAG	GGCCCCCTGCTG	CTGGTCTCAGGATGGG	CTCTGGGGG	489	
Db								
QY	490	CTCTGCTGTT	CATCAAGACAC	CCCCCTCTG	CCCCAGCTCACC	ACATCTTCATC	549	
Db								
QY	490	CTCTGCTGTT	CATCAAGACAC	CCCCCTCTG	CCCCAGCTCACC	ACATCTTCATC	549	
Db								
QY	550	CCAGGACTTGG	GTCCCCCATCTCT	CAGAATCTACA	AGGCCATTGCT	CCCCAGCTGGG	609	
Db								
QY	550	CCAGGACTTGG	GTCCCCCATCTCT	CAGAATCTACA	AGGCCATTGCT	CCCCAGCTGGG	609	
Db								
QY	610	TCAACTCCCC	CCCTGCTTGTG	TGACAGAGGGGGT	GGGACAGGGGGT	GGGGGTTGCC	669	
Db								
QY	610	TCAACTCCCC	CCCTGCTTGTG	TGACAGAGGGGGT	GGGACAGGGGGT	GGGGGTTGCC	669	
Db								

Db	178	TCACCTCC	CCCCCTGCCTCGTTGGTGCACAGGGGGTGGACAGGGG	-GGGGGTCCC	120
QY	670	CCTGTACATACCTGCCATACCAACCCCAAGGTATTAAATCTCGCTGGTTTCTTTTATT			729
Db	119	CCTGTACATACCTGCCATACCAACCCCAAGGTATTAAATCTCGCTGGTTTCTTTTATT			60
QY	730	TTAAATTTTGTGTTTGAATTTTTTAAATAGAAATTTTCATTTTAAGCACA			779
Db	59	TTAAATTTTGTGTTTGAATTTTTTAAATAGAAATTTTCATTTTAAGCACA			10

RESULT 14  
AL578663/C

LOCUS	AL578663	1051 bp	mRNA	linear	EST 16-FEB-2001
DEFINITION	AL578663 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK004YG05 3 prime mRNA sequence.				
ACCESSION	AL578663				
VERSION	AL578663.1 GI:12942954				
KEYWORDS	EST.				
SOURCE	human.				
ORGANISM	Homo sapiens				
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Hominidae; Homo.				
REFERENCE	1 (bases 1 to 1051)				
AUTHORS	Li, W.-B., Gruber, C., Jessee, J. and Pollayes, D.				
TITLE	Full-length cDNA libraries and normalization				
JOURNAL	Unpublished (2001)				
COMMENT	Contact: Genoscope				
	Genoscope - Centre National de Sequencage				
	Bp 191 91006 EVRY cedex - France				
	Email: seqref@genoscope.cns.fr, web : www.genoscope.cns.fr.				

**FEATURES**  
**SOURCE**

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CSODK004YG05"  
 /clone\_lib="LTI\_NFL006\_PL2"  
 /tissue\_type="placenta"  
 /note="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : <http://fulllength.invitrogen.com>"

189 a 311 c 354 g 189 t 8 others

BASE COUNT  
 ORIGIN

[illegible]



GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 14:38:45 ; Search time 158 Seconds  
(without alignments)

11234.997 Million cell updates/sec

Title: US-09-691-220-1

Perfect score: 2086

Sequence: 1 aacgacgagggcgagggg.....cacaaaaa.....2086

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 574371 seqs, 425486471 residues

Total number of hits satisfying chosen parameters: 1148742

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications NA:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US09\_PUBCOMB.seq.\*
- 11: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*
- 12: /cgn2\_6/ptodata/1/pubpna/US10\_PUBCOMB.seq.\*
- 13: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*
- 14: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	1778.8	85.3	2907	10	US-09-954-456-524
2	749.2	35.9	1577	10	US-09-797-727-1
3	701.2	33.6	1375	9	US-10-239-804-7
4	701.2	33.6	1399	9	US-10-239-804-8
5	701.2	33.6	9127	9	US-10-239-804-9
6	701.2	33.6	9151	9	US-10-239-804-10
C 7	591.4	28.4	595	10	US-09-822-830A-614
8	375.6	18.0	411	10	US-09-960-352-5714
C 9	375.6	18.0	417	10	US-09-960-352-5714
10	363.6	17.4	429	10	US-09-960-352-12806
11	362.6	17.4	393	10	US-09-960-352-31
C 12	324.6	15.6	456	10	US-09-833-381-92
13	282.4	13.5	416	10	US-09-960-352-2138
14	178.6	8.6	2288	10	US-09-962-436-567
15	176.2	8.4	1123	10	US-09-965-703-26
16	176	8.4	1404	10	US-09-965-703-62
17	173.8	8.3	312	10	US-09-833-381-1197
18	172.8	8.3	987	10	US-09-965-703-21
19	170.4	8.2	848	10	US-09-833-381-1210

20	159.6	7.7	816	9	US-09-921-650-22	Sequence 22, Appl
21	159.6	7.7	816	10	US-09-874-389-22	Sequence 22, Appl
22	159.6	7.7	1443	9	US-09-921-650-37	Sequence 37, Appl
23	151.2	7.2	1594	10	US-09-964-824A-559	Sequence 559, App
24	150.8	7.2	1934	10	US-09-909-446-1	Sequence 1, Appl
25	150.8	7.2	1934	10	US-09-909-325-1	Sequence 1, Appl
26	150.8	7.2	1934	10	US-09-909-326-1	Sequence 1, Appl
27	149.8	7.2	1528	10	US-09-962-832-109	Sequence 109, App
28	149.8	7.2	1528	10	US-09-880-107-3328	Sequence 3328, App
29	149	7.1	2626	12	US-10-002-600-45	Sequence 45, Appl
30	149	7.1	2670	12	US-10-002-600-44	Sequence 44, Appl
31	147.6	7.1	2481	10	US-09-853-386-89	Sequence 89, Appl
32	146.6	7.0	1679	12	US-10-044-090-316	Sequence 316, App
33	146	7.0	1841	12	US-10-013-823-1	Sequence 1, Appl
34	145.2	7.0	1759	10	US-09-962-436-300	Sequence 300, App
35	138.4	6.6	2880	12	US-10-044-090-361	Sequence 361, App
36	136.4	6.5	3427	10	US-09-853-386-4	Sequence 4, Appl
37	135.8	6.5	3421	10	US-09-853-386-14	Sequence 14, Appl
38	135.2	6.5	455	10	US-09-867-701-4261	Sequence 4261, App
39	134.8	6.5	2469	10	US-09-853-386-90	Sequence 90, Appl
40	132.4	6.3	2218	10	US-09-962-436-268	Sequence 268, App
41	129.8	6.2	1622	10	US-09-925-297-176	Sequence 176, App
42	127.6	6.1	2511	10	US-09-919-497-36	Sequence 36, Appl
43	122.2	5.9	2066	10	US-09-895-840-1	Sequence 1, Appl
44	120.6	5.8	367	10	US-09-960-352-1802	Sequence 1802, App
45	113.4	5.4	850	10	US-09-965-703-28	Sequence 28, Appl

ALIGNMENTS

RESULT 1

US-09-954-456-524  
; Sequence 524, Application US/09954456  
; Patent No. US20020115057A1  
; GENERAL INFORMATION:  
; APPLICANT: Young, Paul  
; TITLE OF INVENTION: Process for Identifying Anti-Cancer Therapeutic Agents Using  
; FILE OF INVENTION: Sets  
; FILE REFERENCE: 689290-76  
; CURRENT APPLICATION NUMBER: US/09/954,456  
; PRIOR FILING DATE: 2001-09-18  
; PRIOR APPLICATION NUMBER: US/60/233,617  
; PRIOR FILING DATE: 2000-09-18  
; PRIOR APPLICATION NUMBER: US/60/234,052  
; PRIOR FILING DATE: 2000-09-20  
; PRIOR APPLICATION NUMBER: US/60/234,923  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,134  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,637  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,638  
; PRIOR FILING DATE: 2000-09-26  
; PRIOR APPLICATION NUMBER: US/60/235,711  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,720  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,840  
; PRIOR FILING DATE: 2000-09-27  
; PRIOR APPLICATION NUMBER: US/60/235,863  
; PRIOR FILING DATE: 2000-09-27  
; NUMBER OF SEQ ID NOS: 2276  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 524  
; LENGTH: 2907  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-954-456-524

Query Match 85.3%; Score 1778.8; DB 10; Length 2907;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 1786; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY	288	CTCCATTGAGACCCAGACAGCAGCTTCTGAAGAGATAGTGCACGCCCTCCCTCGCCACC	347
DB	279	CACCATTGAGACCCAGACAGCAGCTTCTGAAGAGATAGTGCACGCCCTCCCTCGCCACC	338
QY	348	CCCTCTACCCCGCATCTACAAGCCTTGGTTTGTCTGCAGGACAAAGTCTCAGGCTACCA	407
DB	339	CCCTCTACCCCGCATCTACAAGCCTTGGTTTGTCTGCAGGACAAAGTCTCAGGCTACCA	398
QY	408	CTATTGGGCTCAGCGCTCTGAGGGCTGCAAGGGCTTCTCCGCCGAGCATTCAGAGAA	467
DB	399	CTATTGGGCTCAGCGCTCTGAGGGCTGCAAGGGCTTCTCCGCCGAGCATTCAGAGAA	458
QY	468	CATGTGTACACGTGTCAACCGGGAAGAAGTGCATCATCAAGAGTGCACCGGAACC	527
DB	459	CATGTGTACACGTGTCAACCGGGAAGAAGTGCATCATCAAGAGTGCACCGGAACC	518
QY	528	CTGCCAGTACTGCCGACTGCNAAAGTCTTTGAAGTGGGCATGTCCAAGGAGTCTGTGAG	587
DB	519	CTGCCAGTACTGCCGACTGCNAAAGTCTTTGAAGTGGGCATGTCCAAGGAGTCTGTGAG	578
QY	588	AAAGCAGCGAAACAAGAAAGAGAGAGTGCATTCAGAAAGTGCCTCAAGCGCACCAGAACCTT	707
DB	579	AAAGCAGCGAAACAAGAAAGAGAGAGTGCCTCAAGCGCACCAGAACCTT	698
QY	648	GCTGACCCGAGGTGGGGAGTCTCATTCAGAAAGTGCCTCAAGCGCACCAGAACCTT	767
DB	639	GCTGACCCGAGGTGGGGAGTCTCATTCAGAAAGTGCCTCAAGCGCACCAGAACCTT	758
QY	708	CCCTGCCCTTGCAGGTGGGCAAAATACACTAGAACAAACAGCTCAGAACAAAGTGTCTC	827
DB	699	CCCTGCCCTTGCAGGTGGGCAAAATACACTAGAACAAACAGCTCAGAACAAAGTGTCTC	818
QY	768	TCTGGACATTGACCTCTGGGACAAAGTTTCAGTGAACCTCCACCAATCGCGACAGATCAC	887
DB	759	TCTGGACATTGACCTCTGGGACAAAGTTTCAGTGAACCTCCACCAATCGCGACAGATCAC	878
QY	828	TGTGGAGTTCCCAACAGCTGCCCGGCTTCAACACCTTCACCACTCGCGACAGATCAC	947
DB	819	TGTGGAGTTCCCAACAGCTGCCCGGCTTCAACACCTTCACCACTCGCGACAGATCAC	938
QY	888	CTCTCTCAAGGCTGCCTGGACATCTGATCTCGGGATCTGCACCGGATACAGCC	1007
DB	879	CTCTCTCAAGGCTGCCTGGACATCTGATCTCGGGATCTGCACCGGATACAGCC	998
QY	948	CGAGCAGGACACCATGACCTTCTCGGACGGCTGACCTGAAACGGACCCAGATGCACAA	1067
DB	939	CGAGCAGGACACCATGACCTTCTCGGACGGCTGACCTGAAACGGACCCAGATGCACAA	1058
QY	1008	CGCTGGCTTGGCCCCCTCACCGACTTGTCTTTTGCCTTCGCCAACAGCTGCTGCCCT	1127
DB	999	CGCTGGCTTGGCCCCCTCACCGACTTGTCTTTTGCCTTCGCCAACAGCTGCTGCCCT	1118
QY	1068	GGAGATGGATGATGCGGAGACGGGCTGCTCAGCGCCATCTGCCTCATCTCGGAGACCG	1187
DB	1059	GGAGATGGATGATGCGGAGACGGGCTGCTCAGCGCCATCTGCCTCATCTCGGAGACCG	1178
QY	1128	CCAGGACCTGGAGACGCGGACCGGCTGGACATGCTGCAGGACCGGCTGCTGGAGGCGCT	1247
DB	1119	CCAGGACCTGGAGACGCGGACCGGCTGGACATGCTGCAGGACCGGCTGCTGGAGGCGCT	1238
QY	1188	AAAGGTCTACGTGCGGAAGCGGAGCCAGCGCCCCCACATGTTCCCAAGATGCTTAAT	1307
DB	1179	AAAGGTCTACGTGCGGAAGCGGAGCCAGCGCCCCCACATGTTCCCAAGATGCTTAAT	1298
QY	1248	GAAGATTACTGCTGCGAAGCATCAGCGCAAGGGGCTGAGCGGATGATCAGCTGAA	1367
DB	1239	GAAGATTACTGCTGCGAAGCATCAGCGCAAGGGGCTGAGCGGATGATCAGCTGAA	1358
QY	1308	GATGGAGATCCCGGGCTCCATCCGCCCTCTCATCCAGAAATGTTGAGAACTCAGAGG	
DB	1299	GATGGAGATCCCGGGCTCCATCCCGCCCTCTCATCCAGAAATGTTGAGAACTCAGAGG	

```

RESULT 2
US-09-797-727-1
; Sequence 1, Application US/09797727
; Patent No. US20020077457A1
; GENERAL INFORMATION:
; APPLICANT: The Salk Institute for Biological Studies
; APPLICANT: TAKAKU, Fumimaro
; TITLE OF INVENTION: GAMMA RETINOIC ACID RECEPTOR
; FILE REFERENCE: SALK1150-3
; CURRENT APPLICATION NUMBER: US/09/797,727
; CURRENT FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 08/486,325
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/100,039
; PRIOR FILING DATE: 1993-07-30
; PRIOR APPLICATION NUMBER: PCT/US90/03564
; PRIOR FILING DATE: 1990-06-22
; PRIOR APPLICATION NUMBER: US 07/370,407
; PRIOR FILING DATE: 1989-06-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 1
; LENGTH: 1577
; TYPE: DNA

```





Db 530 GAGATGACACGGAGCTAGACGACCTCAGTGAAGATCGGAAAGCCACAGAAACC 589  
Qy 706 TTCCCTGCTCTGCGAGCTGGGCAATACATACAGCAACAGCTCAGAAACACGCTGTC 765  
Db 590 TTTCCTCTACTCTGCGAGCTGGTAAATACACAGAAATTCAGCGCTGACCCCGGTC 649  
Qy 766 TCTCTGGACATTCAGCTCTGGGACAAAGTTTCAAGTAACTTCCACCAAGTGCATTAAG 825  
Db 650 CGATTGGACTTGGGCTCTGGGCAAAATTCAGTGAAGTGGCCACCAAGTGCATTAAG 709  
Qy 826 ACTGTGGAGTTCCGACAGAGCTGCGCGGCTTCAACACCTCACCATCGCGGACCATC 885  
Db 710 ATCGTGGAGTTCCGACAGCTGCGCGGCTTCAAGGTCTGACCATCGAGACCATC 769  
Qy 886 ACCCTCTCAAGCTGCTGCTGCGGACATCTGATCTGCGGATCTCAGCGGATCTCAGCGGATACAG 945  
Db 770 ACCCTGCTCAAGCGGCTGCTGGATATCTTGATCTCAGAAATTCACAGGTATACC 829  
Qy 946 CCGAGAGGACACCATGACCTTCTCGGAGCGGTGACCTGAAACCGGACCATGCAC 1005  
Db 830 CCAGAGACAGACCATGACCTTCTGATGGCTTACACTAAATCGAACTCAGATGCAC 889  
Qy 1006 AACGCTGGCTTGGGCCCCCTCACGAGCTGCTTTTGGCTTGGCCACCAAGCTGCTGCC 1065  
Db 890 AATGCTGGCTTGGGCTCTGACTGACCTTGTGTTCACCTTGTCCAAACCAAGCTCTGCT 949  
Qy 1066 CTGGAGATGATGATGGGAGAGCGGGCTGCTAGCGCCATCTGCCATCTGCTTAACTCTGGAGAC 1125  
Db 950 TTGGAATGATGACAGACAGAAACAGGCTTCTAGTCCCATCTGTTAACTCTGGAGAC 1009  
Qy 1126 CGCCAGAGCTGGAGAGCGGAGCGGCTGATGATGTCAGGAGCGGCTGCTGGAGCG 1185  
Db 1010 CGCCAGAGCTTGGAGAAACCAAAAGTAGAACAGCTCCAAAGAACCACTGCTGGAAGCA 1069  
Qy 1186 CTAAGGTCTAGCTGGGAGAGCGGAGCGGCGCCAGCGGCTGATGTCGCAAGTCTA 1245  
Db 1070 CTAAGATTTACATAGAAACAGAGCGGCTGATGATGTCGCAAGCTTCACTGTTTCCAAAGATCTTA 1129  
Qy 1246 ATGAAGATTTACTGACCTGCGAAGCATCAGGCGCCAGGCGGCTGAGCGGCTGATCAGCGTG 1305  
Db 1130 ATGAATATCAGATCTCCGAGCATCAGCGGAAAGTGGCGAAGCTGTAATACCTTG 1189  
Qy 1306 AAGATGAGATCCGGGCTCAGTCCGCTCTCATCAGGAAATTTGGAGAACTCAGAG 1365  
Db 1190 AAAATGGAATTTCTGATCAATGCCACCTCTCATTCAGGAAATGCTGGAGATTTGAA 1249  
Qy 1366 GGCTGGACACTCTGAGC 1383  
Db 1250 GGACATGAACCTTGACC 1267

RESULT 4

US-10-239-804-8  
; Sequence 8, Application US/10239804  
; Publication No. US20030053991A1  
; GENERAL INFORMATION:  
; APPLICANT: Oxford Biomedica (UK) Limited  
; APPLICANT: Kingsman, Alan J  
; APPLICANT: Maden, Malcolm  
; APPLICANT: Corcoran, Jonathan PT  
; TITLE OF INVENTION: Factor  
; FILE REFERENCE: P009156WOCTH  
; CURRENT APPLICATION NUMBER: US/10/239,804  
; CURRENT FILING DATE: 2002-09-23  
; PRIOR APPLICATION NUMBER: PCT/GB00/01211  
; PRIOR FILING DATE: 2000-03-30  
; PRIOR APPLICATION NUMBER: GB 0024300.6  
; PRIOR FILING DATE: 2000-10-04  
; NUMBER OF SEQ ID NOS: 73  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 8  
; LENGTH: 1399.

; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: FLAG RARbeta2  
; OTHER INFORMATION: PCR product  
US-10-239-804-8  
Query Match 33.6%; Score 701.2; DB 9; Length 1399;  
Best Local Similarity 77.4%; Pred. No. 1.6e-169;  
Matches 850; Conservative 0; Mismatches 248; Indels 0; Gaps 0;  
Qy 286 CACTCCATTGAGACCCAGAGCAGCTTCTGAAGAGATAGTCCCGAGCCCTCCTCGCA 345  
Db 194 CAATCCATCGAGACAGAGTACCAGCTCTGAGGAGCTCGTCCCGAGCCCATCTCCA 253  
Qy 346 CCCCCTCTACCCCGCATCTACAAGCCTTTGCTTCTCTGAGGACAAAGTCTCAGGCTAC 405  
Db 254 CTTCTCTCTCTCGGTGTACAAGCCCTGCTTCGTTTCCAGGACAAAGTCTCAGGCTAC 313  
Qy 406 CACTATGGGTGACGCGCTGTGAGGGTGCAGGGCTTCTTCCCGCGGAGCATCCAGAG 465  
Db 314 CACTATGGGTGACGCGCTGTGAGGGTGCAGGGCTTCTTCCCGAGAAATTCAGAAG 373  
Qy 466 AACATGGTGTACAGCTGTACCGGGACAAAGATGCTCATCATCAACAAAGGTGACCCGGAAC 525  
Db 374 AACATGATCTACACTTGCATCGAGATAAGAACTCGGTCATTAAACAGGTCTACTAGGAAC 433  
Qy 526 CCGTGGCAGTACTGCCGACTGCAGAAAGTCTTGAAGTGGGCTGTCCTCAAGGAGTCTGTG 585  
Db 434 CGATGCGCAGTACTGCCGCTGCAGAAAGTCTTGAAGTGGGCTGTCCTCAAGGAGTCTGT 493  
Qy 586 AGAAAGACCCGAAACAGAAAGAGAGGTGCCCAAGCGGAGTCTCTGAGAGCTAC 645  
Db 494 AGGAATGACAGGAAACAGAAAGAGAGGCTTCAAAAGAGGAAATTCACAGAGAGCTAT 553  
Qy 646 ACCTGCGCGGAGGTGGGAGCTCATTGAGAAGTGGCGCAAGCGCAGCAGCAAGAAC 705  
Db 554 GAGATGACAGCGGAGCTAGACGACCTCTAGAGAGATCCGGAAGAGCCACCGAGNAACC 613  
Qy 706 TTCCCTGCTCTGCCAGCTGGGCAATATACACTAGCAACAAAGCTCAGAAACAGCTGTC 765  
Db 614 TTTCCTCTCCTGCGAGCTGGTAAATACACACAGCAATTCACAGCGCTGACCCCGGGTC 673  
Qy 766 TCTCTGACATTCAGCTCTGGGACAAAGTTCAGTGAACCTTCCACCAAGTGCATTAAG 825  
Db 674 CGATTGGAGTTGGGCTCTGGGACAAATTCAGTGAAGTGGCGCCCAAGTGCATTAAG 733  
Qy 826 ACTGTGGAGTTGCGCAAGCAGCTGCCGCTTCCACACCTCACCATCGCGGACCATGTC 885  
Db 734 ATCGTGGAGTTGCGCAAGCAGCTGCCGCTTCCACAGCTTCCAGAGCTGACCATGTC 793  
Qy 886 ACCCTCTCAAGGCTGCTGCTGGACATCTCTGATCTCTGCGGATCTGACGCGGTACAG 945  
Db 794 ACCCTGCTCAAGCGGCTGCTGGATATCTTGATCTCAGAAATTTGACAGGTATACC 853  
Qy 946 CCCGAGGAGGACACCATGACCTTCTCGGAGCGGCTGACCTGAAACCGGACCATGAC 1005  
Db 854 CCAGAGCAAGACACCATGACCTTCTCTGATGGCTTACACTAAATCGAACTCAGATGAC 913  
Qy 1006 AACGCTGGCTTGGGCCCCCTCACGAGCTGCTTTTGGCTTGGCCAAACAGCTGCTGCC 1065  
Db 914 AATGCTGGCTTGGGCTCTGACTGACCTTGTGTTACCTTTGCCAAACAGCTCTCGCT 973  
Qy 1066 CTGGAGATGGATGATGCGGAGACGGGCTGCTCAGCGCATCTGCTCATCTCTCGGAGAC 1125  
Db 974 TTGGAATGGATGACACAGAAACAGGCTTCTCAGTGCCTCTGTTATCTGTGGAGAC 1033  
Qy 1126 CGCCAGGACCTGGAGACGCGGAGCGGTGGACATGCTGACAGAGCGGCTGCTGGAGCG 1185  
Db 1034 CGCCAGGACCTTGGAGAACCAACAAAGTAGACAAAGTCCAAAGAACCACTGCTGGAAGCA 1093  
Qy 1186 CTAAGGTCTAGTGGGAGAGGAGCGGCGCCCGCCCATGTTCCCGAGATGCTA 1245  
Db 1186 CTAAGGTCTAGTGGGAGAGGAGCGGCGCCCGCCCATGTTCCCGAGATGCTA 1245





ORGANISM: Artificial Sequence

FEATURE: Description of Artificial Sequence:

OTHER INFORMATION: PONY-FLAG-RARBeta2 vector genome plasmid

US-10-239-804-10

Query Match 33.6%; Score 701.2; DB 9; Length 9151;  
Best Local Similarity 77.4%; Pred. No. 4.6e-169;  
Matches 850; Conservative 0; Mismatches 248; Indels 0; Gaps 0;

QY 286 CACTCCATTGAGACCCAGACGAGCTCTGAGAGATAGTCCCGAGCCCTCCCTCGCCA 345  
DB 3018 CAATCCATCGAGACACAGAGTACCAGCTCTGAGAGCTCGTCCCGAGCCCACTCCCA 3077  
QY 346 CCCCCCTACCCCGCATCTACAAAGCTTGTCTGTGAGGACAAAGTCCCTCAGGCTAC 405  
DB 3078 CTTCCTCTCCTCGGTGTACAAAGCTTGTCTGTGAGGACAAAGTCCCTCAGGCTAC 3137  
QY 406 CACTATGGGTGAGCGCTGTGAGGGCTGCAAGGGCTTCTCCGCGGACGATCCAGAAG 465  
DB 3138 CACTATGGGTGAGCGCTGTGAGGGCTGCAAGGGCTTCTCCGCGGAGATATTCAGAAG 3197  
QY 466 AACATGGTGTACACCTGTACCGGACAAAGTCAATCAACAAAGTACCCGGAAC 525  
DB 3198 AACATGGTGTACACCTGTACCGGACAAAGTCAATCAACAAAGTACCCGGAAC 3257  
QY 526 CACTCCAGTACTGCGAGTGTGAGAGTGTGAGTGGGATGTCCTCAAGGAGTCTGTG 585  
DB 3258 CACTCCAGTACTGCGAGTGTGAGAGTGTGAGTGGGATGTCCTCAAGGAGTCTGTG 3317  
QY 586 AGAAACGACCGAAGAAAGAGAGTGTGCGCAAGCCGAGTGTCTGAGAGCTAC 645  
DB 3318 AGAAACGACCGAAGAAAGAGAGTGTGCGCAAGCCGAGTGTCTGAGAGCTAT 3377  
QY 646 ACGTGTGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 705  
DB 3378 GAGATGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 3437  
QY 706 TTCCCTGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 765  
DB 3438 TTCCCTGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 3497  
QY 766 TCTCTGAGATGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 825  
DB 3498 GAGTGTGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 3557  
QY 826 ACTGTGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAG 885  
DB 3558 ACTGTGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAG 3617  
QY 886 ACCCTCTCAAGGCTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 945  
DB 3618 ACCCTCTCAAGGCTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 3677  
QY 946 CCGGAGGAGGACACATGACCTTCTCGGACGGGCTGACCTGAAACCGGACCCAGATGAC 1005  
DB 3678 CCGGAGGAGGACACATGACCTTCTCGGACGGGCTGACCTGAAACCGGACCCAGATGAC 3737  
QY 1006 AACGCTGTGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 1065  
DB 3738 AACGCTGTGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 3797  
QY 1066 CTGAGATGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAG 1125  
DB 3798 CTGAGATGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAG 3857  
QY 1126 CGCCAGGAGCTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAG 1185  
DB 3858 CGCCAGGAGCTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAG 3917  
QY 1186 CTAAGGCTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 1245  
DB 3918 CTAAGGCTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 3977

RESULT 7

US-09-822-830A-614/C  
; Sequence 614, Application US/09822830A  
; Patent No. US20020142952A1  
; GENERAL INFORMATION:  
; APPLICANT: Genetics Institute, Inc.  
; APPLICANT: Wong, Gordon G.  
; APPLICANT: Clark, Hilary  
; APPLICANT: Fechtel, Kim  
; APPLICANT: Agostino, Michael J.  
; APPLICANT: Howes, Steven H.  
; APPLICANT: Resnick, Richard J.  
; APPLICANT: Gulukota, Kamalakar  
; APPLICANT: Graham, James R.  
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING NOVEL SECRETED PROTEINS  
; FILE REFERENCE: GIN 6402  
; CURRENT APPLICATION NUMBER: US/09/822,830A  
; CURRENT FILING DATE: 2001-03-29  
; PRIOR APPLICATION NUMBER: 60/195,604  
; PRIOR FILING DATE: 2000-04-06  
; NUMBER OF SEQ ID NOS: 631  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 614  
; LENGTH: 595  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
US-09-822-830A-614

Query Match 28.4%; Score 591.4; DB 10; Length 595;  
Best Local Similarity 99.8%; Pred. No. 1.3e-141;  
Matches 592; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1492 TCCTGTCGCGCCAGCCACATGACACAGCCCTGCGCCCTCGCCCGCCGCTTTCTCT 1551  
DB 593 TCCTGTCGCGCCAGCCACATGACACAGCCCTGCGCCCTCGCCCGCCGCTTTCTCT 534  
QY 1552 GCCTTTCTACCGACCATGTGACCCCGCACAGCCCTGCCCCACCTGCCCTCCCGGCGAG 1611  
DB 533 GCCTTTCTACCGACCATGTGACCCCGCACAGCCCTGCCCCACCTGCCCTCCCGGCGAG 474  
QY 1612 TACTGGGGACCTTCCCTGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1671  
DB 473 TACTGGGGACCTTCCCTGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 414  
QY 1672 CTGGGCGCTCAGTGGAGTGCCTGCTCCACAGCTGGGCTGAGGTGAGGTGAGGTGAGGTGAG 1731  
DB 413 CTGGGCGCTCAGTGGAGTGCCTGCTCCACAGCTGGGCTGAGGTGAGGTGAGGTGAGGTGAG 354  
QY 1732 GGAATGAGTGGGCGCTTGGTCTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 1791  
DB 353 GGAATGAGTGGGCGCTTGGTCTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAGGTGAG 294  
QY 1792 TCAAGACACCTCTCTCCCGAGTCCACATCTTATCACCAGCAACAGCCAGGACTTGG 1851  
DB 293 TCAAGACACCTCTCTCCCGAGTCCACATCTTATCACCAGCAACAGCCAGGACTTGG 234  
QY 1852 CTCCTCCCGAGTCCAGGAGTCCAGGAGTCCAGGAGTCCAGGAGTCCAGGAGTCCAGGAG 1911  
DB 233 CTCCTCCCGAGTCCAGGAGTCCAGGAGTCCAGGAGTCCAGGAGTCCAGGAGTCCAGGAG 174

Qy	1912	CTGCGCTCGGTTGGTGACAGAGGGGTGGACAGGGGGGGGGTCCCCCTGTACATAC	1971
Db	173	CCTGCCTCGGTTGGTGACAGAGGGGTGGACAGGGGGGGTCCCCCTGTACATAC	114
Qy	1972	CTGCGCATACCAACCCAGGTAAATCTCGCTGGTTTCTGTTTATTTTAAATTTTTT	2031
Db	113	CTGCGCATACCAACCCAGGTAAATCTCGCTGGTTTCTGTTTATTTTAAATTTTTT	54
Qy	2032	GTTTTCATTTTTTAAATAGAATTTTCATTTTAAAGCACAAAAA	2084
Db	53	GTTTTCATTTTTTAAATAGAATTTTCATTTTAAAGCACAAAAA	1

RESULT 8  
 US-09-960-352-5714  
 ; Sequence 5714, Application US/09960352  
 ; Patent No. US20020137139A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Warren, Wesley C.  
 ; APPLICANT: Tao, Nengbing  
 ; APPLICANT: Byatt, John C.  
 ; APPLICANT: Mathialagan, Nagappan  
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTIC ACID  
 ; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
 ; FILE REFERENCE: 16511.006/37-21(10298)C  
 ; CURRENT APPLICATION NUMBER: US/09/960.352  
 ; CURRENT FILING DATE: 2001-09-24  
 ; NUMBER OF SEQ ID NOS: 15112  
 ; SEQ ID NO 5714  
 ; LENGTH: 411  
 ; TYPE: DNA  
 ; ORGANISM: Bos taurus  
 ; OTHER INFORMATION: Clone ID: 25-LIB188-001-Q1-E1-G1  
 US-09-960-352-5714

; GENERAL INFORMATION:  
 ; APPLICANT: Warren, Wesley C.  
 ; APPLICANT: Tao, Nengbing  
 ; APPLICANT: Byatt, John C.  
 ; APPLICANT: Mathialagan, Nagappan  
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
 ; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
 ; FILE REFERENCE: 16511.006/37-21(10298)C  
 ; CURRENT APPLICATION NUMBER: US/09/960,352  
 ; CURRENT FILING DATE: 2001-09-24  
 ; NUMBER OF SEQ ID NOS: 15112  
 ; SEQ ID NO 12806  
 ; LENGTH: 417  
 ; TYPE: DNA  
 ; ORGANISM: Bos taurus  
 ; OTHER INFORMATION: Clone ID: 55-LIB188-005-Q1-E1-F4  
 US-09-960-352-12806

Query Match	18.0%;	Score 375.6;	DB 10;	Length 417;
Best Local Similarity	95.3%;	Pred. No. 2e-86;	Mismatches 19;	Indels 0; Gaps 0;
Matches 387;	Conservative 0;			
Qy 536	ACTGCCGACTGCAGAAAGTGGTTTGAAGTGGGCATGTC	CAAGGAGTCTCTGAGAAACGAC	595	
Db 417	ACTCCGGCTGCAGAAAGTGGTTTGAAGTGGGCATGTC	CAAGGAGTCTCTGAGAAATGACC	358	
Qy 596	GAACAAGAAGAAAGAGAGTGC	CCCAAGCCGAGTCTTGAGAGCTACACGCTGACGC	655	
Db 357	GAACAAGAAGAAAGAGAGTGC	CCCAAGCCGAGTCTTGAGAGCTACACGCTGACGC	298	
Qy 656	CGAGGTGGGGAGCTCATTGAGAAGTGC	CAAGCGCACCGAGAACTTCCCTGGCC	715	
Db 297	CGAGGTGGGGAGCTCATTGAGAAGTGC	CAAGCGCACCGAGAACTTCCCTGGCTC	238	
Qy 716	TCTGCAGCTGGGCAAAATACACTACGAAACACAGCTC	AGAAACAGCTCTCTCTGGACA	775	
Db 237	TCTGCAGCTGGGCAAAATACACTACGAAACACAGCTC	AGAAACAGCTCTCTCTGGACA	178	
Qy 776	TTGACCTCTGGACAAAGTTCAGTGAAGTCTCCACCA	AGTGCATTAAGACTGTGGAGT	835	
Db 177	TTGACCTCTGGACAAAGTTCAGTGAAGTCTCCACCA	AGTGCATTAAGACTGTGGAGT	118	
Qy 836	TGCCCAAGCAGCTGCCCGGCTTCCACCACTC	CACCATCGCGACAGATCACCCCTCTCA	895	
Db 117	TGCCCAAGCAGCTGCCCGGCTTCCACCACTC	CACCATCGCGACAGATCACCCCTCTCA	58	
Qy 896	AGGCTGCCTCGGTGACATCCTGATCCTCGGATCTG	CACCGGTA 941		
Db 57	AGGCTGCCTCGGTGATATCCTGATCCTGGGTATCTG	CACCGGTA 12		

RESULT 10  
 US-09-960-352-31  
 ; Sequence 31, Application US/09960352  
 ; Patent No. US20020137139A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Warren, Wesley C.  
 ; APPLICANT: Tao, Nengbing  
 ; APPLICANT: Byatt, John C.  
 ; APPLICANT: Mathialagan, Nagappan  
 ; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
 ; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION  
 ; FILE REFERENCE: 16511.006/37-21(10298)C  
 ; CURRENT APPLICATION NUMBER: US/09/960,352  
 ; CURRENT FILING DATE: 2001-09-24  
 ; NUMBER OF SEQ ID NOS: 15112  
 ; SEQ ID NO 31  
 ; LENGTH: 429  
 ; TYPE: DNA  
 ; ORGANISM: Bos taurus  
 ; OTHER INFORMATION: Clone ID: 01-LIB188-009-Q1-E1-A1  
 US-09-960-352-31



US-09-960-352-2138  
; Sequence 2138, Application US/09960352  
; Patent No. US20020137139A1  
; GENERAL INFORMATION:  
; APPLICANT: Warren, Wesley C.  
; APPLICANT: Tao, Nengbing  
; APPLICANT: Byatt, John C.  
; APPLICANT: Mathalagan, Nagappan  
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND  
; FILE OF INVENTION: MUSCLE AND FAT DEPOSITION  
; FILE REFERENCE: 16511.006/37-21(10298)C  
; CURRENT APPLICATION NUMBER: US/09/960,352  
; CURRENT FILING DATE: 2001-09-24  
; NUMBER OF SEQ ID NOS: 15112  
; SEQ ID NO 2138  
; LENGTH: 416  
; TYPE: DNA  
; ORGANISM: Bos taurus  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (76)  
; OTHER INFORMATION: unsure at all n locations  
; OTHER INFORMATION: Clone ID: 10-LIB188-020-Q1-E1-C5  
US-09-960-352-2138

Query Match 13.5%; Score 282.4; DB 10; Length 416;  
Best Local Similarity 94.3%; Pred. No. 1.4e-62;  
Matches 314; Conservative 0; Mismatches 17; Indels 2; Gaps 2;

QY	613	GAGGTGCCCAAGCCCGAGTCTCTGAGAGCTACACGCTGA-CGCGGAGGTGGGGAGCT	671
Db	12	GGGATACCAAGCACGAGTCTCCGAGAGCTACACGTTGACCCCGGAGGTGGGGAGCT	71
QY	672	CAT-TGAGAAGGTGGCGAAAGCGCACGAGGAACCTTCCCTGCCCTCTGCCAGCTGGCA	730
Db	72	CATCNGAAGGTGGCGAAAGCGCATCAGAGAGACCTTCCCTGCTCTCTGCCAACTGGCA	131
QY	731	AATPACTAGCAACAACAGCTCAGAACACGCTGCTCTCTTGGACATTGACCTCTGGGACA	790
Db	132	AATPACTAGCAACAACAGCTCAGAACAGCGTGCTCTCTTGGACATTGACCTCTGGGACA	191
QY	791	AGTTCAGTGAACCTCTCCACCAAGTGATCATTAAGACTGTGGAGTTCGCCAACAGCTGC	850
Db	192	AGTTCAGTGAACCTCTCCACCAAGTGATCATTAAGACTGTGGAGTTCGCCAACAGCTGC	251
QY	851	CCGGCTTACACACCTTACCATCGCGAGACAGATCACCTCTCAAGGCTGCCTGGCTGG	910
Db	252	CCGGCTTACACACCTTACCATTGGCGAGCAGATCACCTCTCAAGGCTGCCTGGCTGG	311
QY	911	ACATCTGATCTCTCGGATCTGCACGCGGTACA	943
Db	312	ATATCTGATCTCTCGGATCTGCACGCGGTACA	344

RESULT 14  
US-09-962-436-567  
; Sequence 567, Application US/09962436  
; Patent No. US20020081301A1  
; GENERAL INFORMATION:  
; APPLICANT: Soppet, Daniel  
; TITLE OF INVENTION: Cancer Gene Determination and Therapeutic Screening Using Sign  
; FILE OF INVENTION: Sets  
; FILE REFERENCE: 689290-75  
; CURRENT APPLICATION NUMBER: US/09/962,436  
; CURRENT FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/60/235,082  
; PRIOR FILING DATE: 2000-09-25  
; PRIOR APPLICATION NUMBER: US/60/234,924  
; PRIOR FILING DATE: 2000-09-25  
; NUMBER OF SEQ ID NOS: 568  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 567  
; LENGTH: 2288

Db 1570 GATGAAGTGAAGTCTGACCTCCGATCGGGCTGCCACGCCGCTCTCTCCACAT 1629  
QY 1305 GAAGATGGAGATCC---CGGCTCCATCGCGCTCTCATCCAGGAAATGTTGGAGAACTC 1361  
Db 1630 GAAAGTCCAGTGGCCCCACCGAACTCTTCCGCCCACTCTTCTCGAGGCTTTTGAGGATCA 1689  
QY 1362 AGAGGGCTGGACACTCTGAGCGGACCGCGGGGGTGGGGGGCGGAGCGG 1414  
Db 1690 GGAAGTCTAAGCCTCAGCGCGCCAGAGGGTGTGCGGAGCTGGTGGGAGGAG 1742

RESULT 15

US-09-965-703-26  
; Sequence 26, Application US/09965703  
; Patent No. US20020119521A1  
; GENERAL INFORMATION:  
; APPLICANT: Rohm and Haas Company  
; APPLICANT: Palli, Subba Reddy  
; APPLICANT: Kapitskaya, Marianna Zinovjevna  
; APPLICANT: Cress, Dean Ervin  
; TITLE OF INVENTION: No. US20020119521A1el Ecdysone Receptor-Based Inducible Gene Expr  
; FILE OF INVENTION: A01020B  
; CURRENT APPLICATION NUMBER: US/09/965.703  
; CURRENT FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: 60/191,355  
; PRIOR FILING DATE: 2000-03-22  
; PRIOR APPLICATION NUMBER: 60/269,799  
; PRIOR FILING DATE: 2001-02-20  
; PRIOR APPLICATION NUMBER: PCT/US01/09050  
; PRIOR FILING DATE: 2001-03-21  
; NUMBER OF SEQ ID NOS: 75  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 26  
; LENGTH: 1123  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; NAME/KEY: misc\_feature  
; OTHER INFORMATION: No. US20020119521A1el Sequence  
US-09-965-703-26

Query Match 8.4%; Score 176.2; DB 10; Length 1123;  
Best Local Similarity 51.2%; Pred. No. 3.9e-35;  
Matches 505; Conservative 0; Mismatches 458; Indels 24; Gaps 3;

QY 373 TGCCTTTGCTCTCAGGACAACTCTCAGGCTACCACTATGGGGTCAGCGCCTGTGAGGCG 432  
Db 1 TGCGCCATCTCGGGGACCGCTCTCAGGCAAGCACTATGGAGTGTACAGCTGCGAGGGG 60  
QY 433 TGAAGGCTTCTTCCGCGCAGCATCCAGAAAGACATGGTGTACACGTGTACCGGGAC 492  
Db 61 TGAAGGCTTCTTCAAGCGGACGGTGGCGAAGGACCTGACCTACACCTGCGCGGACAAAC 120  
QY 493 AAGAACTGCATCAACAAGGTGACCGGAAACCCCTGCCAGTACTGCCAGTGTGCGAGTGCAGAG 552  
Db 121 AAGGACTGCCTGATTGACAAGCGGACGCGGAAACCGGTGCCAGTACTGCCGTACCAAGAG 180  
QY 553 TGCCTTTGAAGTGGGATGTCGAAGAGTCTGTGAGAAACGACCCGAAACAAAGAAAGAG 612  
Db 181 TGCCTGGCCATGGGCATGAAGCGGGAAGCCGTGCAAGGAGGAGCG---GCAGCGTGGCAAG 237  
QY 613 GAGGTGCCAAGCCGAGTGTCTGAGAGCTACAGCTGACGCGCGGAGGTGGGGGAGCTC 672  
Db 238 GACCGGAACGAGAATGAGGTGGAGTCGACCAAGCAGCGCCAAACGAGGACATGCCGTGGAG 297  
QY 673 ATTGAGAAGTGGCGAAAGCGCACAGGAAACCTTCTCCCTGCGCTGTGCCAGTGGGCAAA 732  
Db 298 AGGATCTGGAGGTGAGCTGGCGTGGAGCCCAAGACCGAGACCTAGCTGGAGGCAAAAC 357  
QY 733 TACACTAGAAACACAGCTCAGAAACAACTGTCTCTCTGGACATTGACCTCTGGGACAAAG 792  
Db 358 ATGGGGCTGAACCCCAAGCTC-----GCCGAAGCAACCCCTGTCAACCAAC 399

QY 793 TTCACTGAACCTCTCCACCAAGTGCATCTTAAGACTGTGAGTTCCGCAAGCAGCTGCC 852  
Db 400 ATTTGCCAAGACGACGACAAACAGCTTTTACCTCGTGGTGGCGCAAGCGGATCCCA 459  
QY 853 GGCCTTACACCACCTCACCATCGCCGACACAGATCACCTCTCTCAAGGCTGCCTGCTGGAC 912  
Db 460 CACTTCTCAGAGCTGCCCTGCGACGACAGGTCTATCTCTGCGGCGAGGCTGGAATGAG 519  
QY 913 ATCTGTATCTCGGATGTGACGGGTACACGCGGAGACCGGACGACACCATGACCTTCTCG 972  
Db 520 CTGCTCATCGCTCTCTCTCCACCGCTCCATCGCGTGAAGGACGCGGATCTCTCTGGCC 579  
QY 973 GACGGCTGACCTGAAACGGGACCCAGATGCACACGCTGGCTTCGCGCCCTCACCGAC 1032  
Db 580 ACCGGCTGCACGTCCACCGGAACAGCGCCACAGCGAGGGTGGCGGCCATCTTTTGGC 639  
QY 1033 CTGGTCTT---TGCCTTGCCTAACCAAGCTGTCTGCCCTGGAGATGGATGATGCGGAGACG 1089  
Db 640 AGGGTCTCAGCGAGCTTGTGTCCAAGATGCGGACATGCAGATGCACAAGACGAGCTG 699  
QY 1090 GGGCTGCTCAGCGCCATCTGCTCATCTCGGGAGACCGCCAGACCTGGAGACGCGGAC 1149  
Db 700 GGTCTGCTGCGCGCATCTCTCTTTAAACCTGACTCCAAGGGGCTCTCGAACCCTGGCC 759  
QY 1150 CGGGTGGACATGCTCAGGAGCGCTGCTGGAGGCGCTAAAGTCTACGTGCGGAAGCGG 1209  
Db 760 GAGGTGGAGGCTGAGGGAAGGTCTATGCGTCTTGGAGGCTTCTGGAAGCAAG 819  
QY 1210 AGGCGGACGCGCCACATGTTCCCAAGATGCTAAATGAAGATTACTGACCTGCGAAGC 1269  
Db 820 TACCCAGAGCAGCGCGGAAGGTTCGCTAAGCTTTGCTCCGCTGCGGCTCTGCGCTCC 879  
QY 1270 ATCAGCGCAAGGGGCTGAGCGGTGATCAGCTGGAAGATGAGATCCCGGCTCCATG 1329  
Db 880 ATCGGGCTCAATGCTGGAACATCTCTTCTTCAAGCTCATCGGGGACACACCAAT 939  
QY 1330 CGGCTCTCATCCAGGAAATGTTGGAG 1356  
Db 940 GACACTTCTTATGAGATGCTGGAG 966

Search completed: March 29, 2003, 18:49:42  
Job time : 201 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 09:48:44 ; Search time 2310 Seconds

(without alignments)

14625.015 Million cell updates/sec

Title: US-09-691-220-1

Perfect score: 2086

Sequence: 1 aacgacgagggcgagggg.....cacaaaaa 2086

Scoring table: IDENTITY\_NUC

Searched: Gapop 10.0 , Gapext 1.0

16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

- 1: em\_estba:\*
- 2: em\_esthum:\*
- 3: em\_estin:\*
- 4: em\_estnu:\*
- 5: em\_estov:\*
- 6: em\_estpl:\*
- 7: em\_estro:\*
- 8: em\_hic:\*
- 9: gb\_estl:\*
- 10: gb\_est2:\*
- 11: gb\_hic:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: em\_estfun:\*
- 16: em\_estom:\*
- 17: gb\_gss:\*
- 18: em\_gss\_hum:\*
- 19: em\_gss\_inv:\*
- 20: em\_gss\_pln:\*
- 21: em\_gss\_vrt:\*
- 22: em\_gss\_fun:\*
- 23: em\_gss\_mam:\*
- 24: em\_gss\_mus:\*
- 25: em\_gss\_other:\*
- 26: em\_gss\_pro:\*
- 27: em\_gss\_rod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	960.8	46.1	1051	9	AL578663
C 2	923.6	44.3	967	9	AL533351
C 3	923.4	44.3	978	9	AL522712
C 4	903.6	43.3	992	9	AL556392
C 5	882.2	42.3	931	9	AL531149
C 6	842.8	40.4	1073	13	BM544324
					BM544324 AGENCOURT

C 7	807.2	38.7	993	9	AL551103
C 8	729.6	35.0	885	14	BQ706025
C 9	678.8	32.5	3141	11	BC030234
C 10	659.8	31.6	702	13	BI827961
C 11	654	31.4	634	14	BQ044936
C 12	648.4	31.1	1091	14	BM927280
C 13	645.4	30.9	703	9	AI806984
C 14	639	30.6	1164	12	BG028346
C 15	635	30.4	646	14	BM790595
C 16	633.8	30.4	700	13	BI907041
C 17	623	29.9	692	9	AA812217
C 18	620.2	29.7	812	13	BI914043
C 19	604.2	29.0	861	10	BE547412
C 20	599.4	28.7	655	12	BG818232
C 21	597.8	28.7	626	9	AI963441
C 22	585	28.0	1123	13	BM468687
C 23	572.8	27.5	602	10	AW118835
C 24	569.8	27.3	764	9	AA790328
C 25	566.8	27.2	670	12	BG337888
C 26	566.8	27.2	1079	14	BM909949
C 27	562.8	27.0	585	13	BI770450
C 28	560	26.8	560	14	BM817785
C 29	559.2	26.8	596	9	AI962618
C 30	559.2	26.8	982	12	BE794377
C 31	550.4	26.4	866	12	BF182871
C 32	546.2	26.2	552	9	AI380422
C 33	540.4	25.9	1034	12	BE871903
C 34	529	25.4	743	13	BI830768
C 35	513	24.6	885	13	BI663939
C 36	504.8	24.2	911	9	AL526754
C 37	502.8	24.1	583	9	AI401587
C 38	495	23.7	954	14	BQ951520
C 39	494.4	23.7	1067	13	BM423874
C 40	492	23.6	921	14	BQ714956
C 41	488.2	23.4	564	10	AV609965
C 42	484.6	23.2	782	13	BI856804
C 43	483.4	23.2	1061	14	BQ278797
C 44	483.4	23.2	1160	13	BM554172
C 45	483	23.2	483	9	AI640571

#### ALIGNMENTS

RESULT 1  
AL578663/C  
LOCUS  
DEFINITION  
AL578663 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CS0DK004YG05 3  
prime, mRNA sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT  
FEATURES  
source

1051 bp mRNA linear EST 16-FEB-2001  
AL578663 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CS0DK004YG05 3  
prime, mRNA sequence.  
AL578663.1 GI:12942954  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 1051)  
Li.W.B., Gruber,C., Jessee,J. and Polayes,D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
Location/Qualifiers  
1..1051  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DK004YG05"  
/clone\_lib="LTI\_NFL006\_PL2"  
/tissue\_type="placenta"  
/note="Vector: PCWSPORT 6; Site\_1: NotI; 1st strand cDNA  
was primed with a NotI-oligo(dT) primer. Five prime end



enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 189 a 311 c 354 g 189 t 8 others  
ORIGIN

Query Match 46.1%; Score 960.8; DB 9; Length 1051;  
Best Local Similarity 97.2%; Pred. No. 8.5e-127;  
Matches 1023; Conservative 7; Mismatches 17; Indels 6; Gaps 5;

QY 992 GGACCCAGATGCACACGCTGCTCGGCCCTCACCACCTGCTTTGGCTTCGCCA 1051  
DB 1051 GGACCCAGATGCACACGCTGCGC-TYGGCCCTTCACCGACTTGCTTTGGCTTCGCCA 993  
QY 1052 ACCAGCTGCTGCCCTGGAGATGGATGATGCGGAGACGGGGTGTCTACGCGCATCTGCC 1111  
DB 992 ACCAGCTGCTGCCCTGGAGATGGATGATGCGGAGACGGGGTGTCTACGCGCATCTGCC 993  
QY 1112 TCATCTGGGAGACCGCAGGACCTGGAGCA-GCCGGACCGGGTGGACATGCTGCAGGAG 1170  
DB 932 TCATCTGGGAGACCGCAGGACCTGGAGCAGCAGCGCCGGTGGACATGCTGCAGGAG 873  
QY 1171 CGCTGCTGGAGGCGCT-AAAGGTCTACGTGCGGAAGCGGAGCGCCAGCGCCCGC--AC 1227  
DB 872 CGCTGCTGGAGGCGCTAAAGGTCTACGTGCGGAAGCGGAGCGCCAGCGCCCGCCACH 813  
QY 1228 ATGTTCCCAAGATGCTAATGAAGATTACTGACCTGCGAAGCATCAGCGCCAGGGGCT 1287  
DB 812 ATGTTCCCAAGATGCTAATGAAGATTACTGACCTGCGAAGCATCAGCGCCAGGGGCT 753  
QY 1288 GAGCGGTGATCAGCCTGAAGATGAGATGCCGGCTCCATGCCGCTCTCATCCAGGAA 1347  
DB 752 GAGCGGTGATCAGCCTGAAGATGAGATGCCGGCTCCATGCCGCTCTCATCCAGGAA 693  
QY 1348 ATGTTGGAGAACTCAGAGGCTGACACTCTGACCGACACCGCGGGGTGGGGGCGG 1407  
DB 692 ATGTTGGAGAACTCAGAGGCTGACACTCTGACCGACACCGCGGGGTGGGGGCGG 633  
QY 1408 CACGGGTGCGCTCCCGCCCGCAGGACGCTGTAGCCCGCAGCTCAGCCCGCAGCTCC 1467  
DB 632 CACGGGTGCGCTGCGCCCGCCCGCAGGACGCTGTAGCCCGCAGCTCAGCCCGCAGCTCC 573  
QY 1468 AACAGACAGCGCGCCACCCACTCCCGGTGACCGCCCGCAGCCACATGGACACAGCCCT 1527  
DB 572 AACAGACAGCGCGCCACCCACTCCCGGTGACCGCCCGCAGCCACATGGACACAGCCCT 513  
QY 1528 CGCCCTCCCGCCCGCTTTCTCTGCTTTCTACCGACATGTGACCCGCGACAGCCCT 1587  
DB 512 CGCCCTCCCGCCCGCTTTCTCTGCTTTCTACCGACATGTGACCCGCGACAGCCCT 453  
QY 1588 GCCCCACTGCTCCCGCGGAGTCTGCGGACCTTCCTCGGGGAGCGGGAGGAGGA 1647  
DB 452 GCCCCANCTG-CTTCCCGGCGAGTCTGGGAGCCTTCCTCGGGGAGCGGGAGGAGGA 394  
QY 1648 GCGAGCACTCTTGGACAGAGGCTGGGCCCTCAGTGACCTGCTGCTCCACAGCCCTG 1707  
DB 393 GCGAGCACTCTTGGACAGAGGCTGGGCCCTCAGTGACCTGCTGCTCCACAGCCCTG 334  
QY 1708 GCGTACCTCAGAGCGCGGAGGAACTGAGTGAGGCCCTTGTCTGGGTCTCAGGA 1767  
DB 333 GCGTACCTCAGAGCGCGGAGGAACTGAGTGAGGCCCTTGTCTGGGTCTCAGGA 274  
QY 1768 TGGTCTCGGGGCTCTGTTTCATCAGACACCCCTCTGCCACCTCACACATCTTCA 1827  
DB 273 TGGTCTCGGGGCTCTGTTTCATCAGACACCCCTCTGCCACCTCACACATCTTCA 214  
QY 1828 TCACAGCAACCGCAGGACTTGGCTCCCGCTCCTCAGAACTCAGAGCCATTCCTCC 1887  
|||||

DB 213 TCACAGCAACGCCAGGACTTGBBTTCCCTCCCTCAGTACTCACAGCCATTGCTCCC 154  
QY 1888 CAGCTGGGAACTCAACCTCCCTCGCTCGTGGTGACAGAGGGGTGGACAGGG 1947  
DB 153 CRGCTGGGAACTCAACCTCCCTCGCTCGTGGTGACAGAGGGGTGGACAGGG 94  
QY 1948 GCGGGGGTTCCTCGTGTACATACCTGCTGCTACCAACCCAGGTATTAATCTCGCTGG 2007  
DB 93 GCGGGGGTTCCTCGTGTAAATAACCTGCTACCAACCCAGGTATTAATCTCGCTGG 34  
QY 2008 TTTGTTTTTATTTTATTTTGTGTTTGTGATT 2040  
DB 33 TTTGTTTTTATTTTATTTTGTGTTTGTGATK 1  
RESULT 2  
AL533351  
LOCUS  
DEFINITION AL533351 LTI\_FL015\_Brn1 Homo sapiens cDNA clone EST 13-FEB-2001  
prime, mRNA sequence.  
ACCESSION AL533351  
VERSION AL533351.1 GI:12796844  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 967)  
AUTHORS Li,W.B., Gruber,C., Jessee,J. and Polayes,D.  
TITLE Full-length cDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
FEATURES  
1..967  
source  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DN004Y122"  
/clone\_lib="LTI\_FL015\_Brn1"  
/sex="male"  
/tissue.type="Adult brain"  
/note="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"  
BASE COUNT 173 a 325 c 286 g 177 t 6 others  
ORIGIN  
Query Match 44.3%; Score 923.6; DB 9; Length 967;  
Best Local Similarity 98.8%; Pred. No. 1.6e-121;  
Matches 956; Conservative 5; Mismatches 4; Indels 3; Gaps 3;  
QY 1083 GGAGAGGGGCTGCTCAGCGCCATCTGCTCCTCATCTCGGAGACCGCCAGACCTTGAGCA 1142  
DB 1 GGAGAGGGGCTGCTCAGCGCCATCTGCTCCTCATCTCGGAGACCGCCAGACCTTGAGCA 60  
QY 1143 GCGGACCGGGTGGACATGCTCAGGAGCGCTGCTGAGGCGCTAAAGCTCTACGTCG 1202  
DB 61 GCGGACCGGGTGGACATGCTCAGGAGCGCTGCTGAGGCGCTAAAGCTCTACGTCG 120  
QY 1203 GAAGGGAGGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGAAGATTACTGACCT 1262  
DB 121 GAAGGGAGGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGAAGATTACTGACCT 180  
QY 1263 GCGAAGCATCAGCGCCAGGGGGTGGAGCGGTGATCATCGCTGAAGATGAGATCCGGG 1322





```

QY 1734 AACTGAGTGAGGCCCTGGTCTCGGTCTCAGGATGGGCTCTGGGGCTCGGTGTTTCATC 1793
Db 308 AACTGAGTGAGGCCCTGGTCTCGGTCTCAGGATGGGCTCTGGGGCTCGGTGTTTCATC 249
QY 1794 AAGACACCCCTGTGCCAGCTCACACATCTTCATCACCAGCAAAAGCCAGGACTTGGGT 1853
Db 248 AAGACACCCCTGTGCCAGCTCACACATCTTCATCACCAGCAAAAGCCAGGACTTGGGT 189
QY 1854 CCCCCTCTCAGAACTCACAAGCCATTGCTCCCAAGCTGGGGAACCTCAACCTCCCCC 1913
Db 188 CCCCCTCTCAGAACTCACAAGCCATTGCTCCCAAGCTGGGGAACCTCAACCTCCCCC 129
QY 1914 TGCCTCGGTGTGACAGAGGGGTGGGACAGGGGGGGGTTCCTCCCTGTATACATACC 1973
Db 128 TSCCTCGGTGTGACAGAGGGGTGGGACAGGGGGGGGTTCCTCCCTGTATACATACC 69
QY 1974 TCCCATACCAACCCAGGTATTAATTCCTCGGTGTTTCTTTTATTTATTTTGT 2033
Db 68 TGCCATACCAAGCAGGTATTAATTCCTCGGTGTTTCTTTTATTTATTTTGT 9
QY 2034 TTTGATTT 2041
Db 8 TTTGATTT 1

RESULT 4
AL556392 992 bp mRNA linear EST 16-FEB-2001
LOCUS
DEFINITION AL556392 LTI_NFL006_PL2 Homo sapiens cDNA clone CS0DK004YG05 5
prime, mRNA sequence.
ACCESSION AL556392
VERSION AL556392.1 GI:12899025
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS Li, W.B., Gruber, C., Jessee, J., and Polayes, D.
TITLE Full-length cDNA libraries and normalization
JOURNAL Unpublished (2001)
COMMENT Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 EVRY cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES
source
1..992
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="CS0DK004YG05"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/note="vector: pcwvSPORT 6; Site_1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched, double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pcwvSPORT 6
vector. Library was normalized. Library was constructed by
Life technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"

BASE COUNT 228 a 319 c 269 g 173 t 3 others
ORIGIN

Query Match 43.3%; Score 903.6; DB 9; Length 992;
Best Local Similarity 98.9%; Pred. No. 1.le-118;
Matches 928; Conservative 1; Mismatches 7; Indels 2; Gaps 2;

QY 288 CTCATTGAGACCCAGACGAGTTCGTGAAGAGATAGTGGCCAGCCCTCCCTCGCCACC 347
Db 35 CACCATTGAGACCCAGACGAGTTCGTGAAGAGATAGTGGCCAGCCCTCCCTCGCCACC 94

```

```

QY 348 CCCTCTACCCCGCATCTACAGCCTTCTGCTTGTCTGTCAGGACAGTCTCTCAGGCTACCA 407
Db 95 CCCTCTACCCCGCATCTACAGCCTTCTGCTTGTCTGTCAGGACAGTCTCTCAGGCTACCA 154
QY 408 CTATGGGGTCAGGCCCTGTGAGGGCTGCAAGGGCTTCTTCGCCCGCAGCATCCAGAGAA 467
Db 155 CTATGGGGTCAGGCCCTGTGAGGGCTGCAAGGGCTTCTTCGCCCGCAGCATCCAGAGAA 214
QY 468 CATGGTGTACACGTGTACCGGGACAAAGAACTGCATCATCAAAAGGTGACCCGGAACCC 527
Db 215 CATGGTGTACACGTGTACCGGGACAAAGAACTGCATCATCAAAAGGTGACCCGGAACCC 274
QY 528 CTCGCCAGTACTCGCCAGTGCAGAAAGTCTTTGAAGTGGGATGTCCCAAGGAGTCTGTGAG 587
Db 275 CTCGCCAGTACTCGCCAGTGCAGAAAGTCTTTGAAGTGGGATGTCCCAAGGAGTCTGTGAG 334
QY 588 AAACGACCCGAAACAAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTCTGAGAGCTACAC 647
Db 335 AAACGACCCGAAACAAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTCTGAGAGCTACAC 394
QY 648 GCTGACCGCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGAAAGCGCACAGGAAACCTT 707
Db 395 GCTGACCGCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGAAAGCGCACAGGAAACCTT 454
QY 708 CCCTGCCCTCTGCCAGCTGGGCAATACACTACGAAACAGCTCAGAAACACCTGTCTC 767
Db 455 CCCTGCCCTCTGCCAGCTGGGCAATACACTACGAAACAGCTCAGAAACACCTGTCTC 514
QY 768 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTCCACCAAGTGCATCATTAAGAC 827
Db 515 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTCCACCAAGTGCATCATTAAGAC 574
QY 828 TGTGGAGTTGCGCAAGCAGCTGCCGGCTTCCACACCTCACCATCGCCGACAGATCAC 887
Db 575 TGTGGAGTTGCGCAAGCAGCTGCCGGCTTCCACACCTCACCATCGCCGACAGATCAC 634
QY 888 CTCTCTCAAGGCTCGCTGTCGACATCTCGGATCTCGGATCTGCGGATCTGCGGATCACGCC 947
Db 635 CTCTCTCAAGGCTCGCTGTCGACATCTCGGATCTCGGATCTGCGGATCTGCGGATCACGCC 694
QY 948 CGAGCAGGACACCATGACCTTCTCGGACGGGCTGACCTGAAACCGGACCCAGATGCACAA 1007
Db 695 CGAGCAGGACACCATGACCTTCTCGGACGGGCTGACCTGAAACCGGACCCAGATGCACAA 754
QY 1008 CGTGTGCTTGGGCCCTTCCACGACCTGTGTTGCTTTCGCCAACAGTGTGTCGCCCT 1067
Db 755 CGTGTGCTTGGGCCCTTCCACGACCTGTGTTGCTTTCGCCAACAGTGTGTCGCCCT 814
QY 1068 GGAGATGGATGATGCGGAGACGGGGCTGCTCAGCGCCATCTGCCTCATCTTCCGGAGACCG 1127
Db 815 GGAGATGGATGATGCGGAGACGGGGCTGCTCAGCGCCATCTGCCTCATCTTCCGGAGACCG 874
QY 1128 CCAGGACCTT-GGAGCAGCCGGACCCGGGTGGACATGCTGAGGAGCGCTGCTGTGAGAGCGC 1186
Db 875 CCAGGACCTTGGAGCAGCCGGGTGGACATGCTGAGGAGCGCTGCTGTGAGAGCGC 934
QY 1187 TAAAGTCTACGTGGGGAAGCGGAGGCCCGACGCCGCC 1224
Db 935 TAAAGTCTACGTGGGGAAGCGGAGGCCCGACGCCGCC 971

RESULT 5
AL531149/c 931 bp mRNA linear EST 13-FEB-2001
LOCUS
DEFINITION AL531149 LTI_NFL001_NBC4 Homo sapiens cDNA clone CS0DM001YN13 3
prime, mRNA sequence.
ACCESSION AL531149
VERSION AL531149.1 GI:12794642
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE 1 (bases 1 to 931)  
 AUTHORS Li, W.B., Gruber, C., Jessee, J. and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 EVRY cedex - France  
 Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
 Location/Qualifiers  
 1..931  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0DM001YN13"  
 /clone\_lib="LRI\_NFL001\_NBC4"  
 /sex="male"  
 /tissue\_type="neuroblastoma cells"  
 /lab\_host="DH10B"  
 /note="Organ: brain; Vector: pCMVSPORT 6; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com"

BASE COUNT 166 a 275 c 316 g 172 t 2 others  
 ORIGIN  
 Query Match 42.3%; Score 882.2; DB 9; Length 931;  
 Best Local Similarity 99.3%; Pred.No. 1.2e-115;  
 Matches 917; Conservative 0; Mismatches 3; Indels 3; Gaps 3;

QY 1121 GAGACCGCCAGGACCTGGAGCAGCGGACCGGGTGGACATCTGCAGGAGCGCTGCTG 1179  
 DB 925 GAGACCGCCAGAGCTGGAGCAGCGGACCGGGTGGACATCTGCAGGAGCGCTGCTG 866  
 QY 1180 GAGGCGCTAAGGCTACGTGGGAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGG 1239  
 DB 865 AAGCGCTAAGGCTACGTGGGAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 806  
 QY 1240 ATGCTAATGAAGATTACTGCTGGAAGCATCAGCGGCGGCGGCGGCGGCGGCGGCGG 1299  
 DB 805 ATGCTAATGAAGATTACTGCTGGAAGCATCAGCGGCGGCGGCGGCGGCGGCGGCGG 746  
 QY 1300 ACGCTGAAGATGGAGATCCCGGGCTCCATCGCGCTCTCATCCAGGAATGTTGGAGAAC 1359  
 DB 745 ACGCTGAAGATGGAGATCCCGGGCTCCATCGCGCTCTCATCCAGGAATGTTGGAGAAC 686  
 QY 1360 TCAGAGGCGCTGGACACTCTGAGCGGACAGCGGGGGGTGGGGGGGCGGAGCGGGGTGGC 1419  
 DB 685 TCAGAGGCGCTGGACACTCTGAGCGGACAGCGGGGGGTGGGGGGGCGGAGCGGGGTGGC 626  
 QY 1420 CTGCGCCCGCCGCGGAGCTGTAGCGCCAGCTCAGCGCCAGCTCCCAACAGAGGACGC 1479  
 DB 625 CTGCGCCCGCCGCGGAGCTGTAGCGCCAGCTCAGCGCCAGCTCCCAACAGAGGACGC 566  
 QY 1480 CCGGCGCCACCCACTCCC - CGTGACCGCCCGCCAGCCACATGGACACAGCCCTCGGCCCTCCGCC 1538  
 DB 565 CCGGCGCCACCCACTCCC GCGGCTGACCGCCAGCCACATGGACACAGCCCTCGGCCCTCCGCC 506  
 QY 1539 CCGGCTTTCTGCTTTCTACCGACCATGTGACCGCCCGCCAGCCCTCGGCCCTCCGCCCTG 1598  
 DB 505 CCGGCTTTCTGCTTTCTACCGACCATGTGACCGCCCGCCAGCCCTCGGCCCTCCGCCCTG 446  
 QY 1599 CCCTCCGGGAGTACTGGGACCTTCCCTGGGGACCGGGAGGAGGAGGAGGAGGAGTCTC 1658  
 DB 445 CCCTCCGGGAGTACTGGGACCTTCCCTGGGGACCGGGAGGAGGAGGAGGAGGAGTCTC 386  
 QY 1659 CTTGGACAGAGGCGCTGGGCGCTCAGTGACATGCGCTGCTCCACAGCCCTGGGCTGACGTCA 1718  
 DB 385 CTTGGACAGAGGCGCTGGGCGCTCAGTGACATGCGCTGCTCCACAGCCCTGGGCTGACGTCA 326

QY 1719 GAGCCGAGCCAGGAACTGAGTGAGGCCCTGCTGCTGGTCTCAGAGTGGGCTCTGGG 1778  
 DB 325 GAGCCGAGCCAGGAACTGAGTGAGGCCCTGCTGCTGGTCTCAGAGTGGGCTCTGGG 266  
 QY 1779 GGCTCGTGTTCATCAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGC-AA 1837  
 DB 265 GGCTCGTGTTCATCAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCTAA 206  
 QY 1838 ACGCCAGGACTTGGCTCCCCCATCTCAGAACTCACAAGCCATTGCTCCCCAGCTGGGGA 1897  
 DB 205 ACGCCAGGACTTGGCTCCCCCATCTCAGAACTCACAAGCCATTGCTCCCCAGCTGGGGA 146  
 QY 1898 ACCTCAACCTCCCTCGCTCGCTGGTGGTGACAGAGGGGTGGGACAGGGCGGGGGTT 1957  
 DB 145 ACCTCAACCTCCCTCGCTCGCTGGTGGTGACAGAGGGGTGGGACAGGGCGGGGGTT 86  
 QY 1958 CCCCCTGTACATACCCTGCCATACCAACCCAGCTATTATTCGCTGGTGTGTTTTT 2017  
 DB 85 CCCCCTGTACATACCCTGCCATACCAACCCAGCTATTATTCGCTGGTGTGTTTTT 26  
 QY 2018 ATTTAATTTTTTTGTTGATT 2040  
 DB 25 ATTTAATTTTTTTGTTGATT 3

RESULT 6  
 BM544324  
 LOCUS BM544324  
 DEFINITION 5', mRNA sequence.  
 ACCESSION BM544324  
 VERSION BM544324.1 GI:18775485  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 1073)  
 AUTHORS NIH-MGC http://mgi.nci.nih.gov/  
 TITLE NIH-MGC  
 JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
 COMMENT Unpublished (1999)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cgapbs-r@mail.nih.gov  
 Tissue Procurement: Invitrogen  
 cDNA Library Preparation: Life Technologies, Inc.  
 DNA Sequencing by: Agencourt Bioscience Corporation  
 Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
 Plate: LLAM12357 row: p column: 20  
 High quality sequence start: 45  
 High quality sequence stop: 712.

FEATURES  
 Location/Qualifiers  
 1..1073  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5587843"  
 /clone\_lib="NIH\_MGC\_125"  
 /lab\_host="DH10B"  
 /note="Organ: ovary (pool of 3); Vector: pCMV-SPORT6; Site 1: EcoRV (destroyed); Site 2: NotI; RNA source pool of three ovaries, from females ranging in age from 38 to 49 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 2.1 kb, insert size range 1-3.5 kb. Library is normalized and enriched for full-length clones and constructed by C. Gruber (Invitrogen). Research Genetics tracking code 036."

BASE COUNT 241 a 335 c 302 g 194 t 1 others  
 ORIGIN

Query Match	40.4%;	Score 842.8;	DB 13;	Length 1073;
Best local Similarity	95.2%;	Pred. No. 3.8e-110;		
Matches 879;	Conservative	0;	Mismatches 43;	Indels 1;
Gaps	1;			
QY	151	GGTCCACCCCTAAATCCCTTCCTACTGTGGATTTTATATACCAGAACCGGGCTGTTTG	210	
DB	3	GGACCGGTCCGGAAATTTTCGGATGTGGTGAATTTTATAACCAAGTATCGGGCTGTTTG	62	
QY	211	CTCCAGAGAAGGGCTCCCGGCCCGGGTCCGTACTCCACCCCGCTCCGAGCTCCCGCTT	270	
DB	63	CTCCAGAGAAGGGCTCCCGGCCCGGGTCCGTACTCCACCCCGCTCCGAGCTCCCGCTT	122	
QY	271	TGGAATGGCTCAACCACTCCATTGAGACCCAGAGCAGCTTCTGAAGAGATAGTGCC	330	
DB	123	TGGAATGGCTCAACCACTCCATTGAGACCCAGAGCAGCTTCTGAAGAGATAGTGCC	182	
QY	331	AGCCCTCCCTCGCCACACCCCTTACCCCGCATCTACAGCCCTTCTGTGTGTCGTCAGGAC	390	
DB	183	AGCCCTCCCTCGCCACACCCCTTACCCCGCATCTACAGCCCTTCTGTGTGTCGTCAGGAC	242	
QY	391	AAGTCTCAGGCTACCACTATATGGGTGAGCGCTGTGAGGGCTCAAGAGGCTTCTTCGCG	450	
DB	243	AAGTCTCAGGCTACCACTATATGGGTGAGCGCTGTGAGGGCTCAAGAGGCTTCTTCGCG	302	
QY	451	CGCAGCTCCAGAGACATGGTGTACAGTGTCCACGGGACAGAACTGCATCATCAAC	510	
DB	303	CGCAGCTCCAGAGACATGGTGTACAGTGTCCACGGGACAGAACTGCATCATCAAC	362	
QY	511	AAGTGTACCCGGAAACCCCTGCGACTGTCCGACTGCGAGAGTCTTGAAGTGGGATG	570	
DB	363	AAGTGTACCCGGAAACCCCTGCGACTGTCCGACTGCGAGAGTCTTGAAGTGGGATG	422	
QY	571	TCCAAGGAGTCTGTGAGAAACGACCCGAAACAGAGAGAGAGGTGCCAAGCCCGAG	630	
DB	423	TCCAAGGAGTCTGTGAGAAACGACCCGAAACAGAGAGAGAGGTGCCAAGCCCGAG	482	
QY	631	TGCTCTGAGAGCTACAGCTGTACGCCGGAGGTGGGGAGCTCATTGAGAAAGTGCACAA	690	
DB	483	TGCTCTGAGAGCTACAGCTGTACGCCGGAGGTGGGGAGCTCATTGAGAAAGTGCACAA	542	
QY	691	CGCACCAGGAAACCTTCCCTGCCCTCTGCGAGTGGGCAATACACTAGCAACACAGC	750	
DB	543	CGCACCAGGAAACCTTCCCTGCCCTCTGCGAGTGGGCAATACACTAGCAACACAGC	602	
QY	751	TCAGAACAGTGCTCTCTGAGACATGACCTCTGGGACAAAGTTCAGTGAACCTCCACC	810	
DB	603	TCAGAACAGTGCTCTCTGAGACATGACCTCTGGGACAAAGTTCAGTGAACCTCCACC	662	
QY	811	AAGTGCATCATTAAGACTGTGGAGTTCGCCAACAGCTGCCCGGCTTACACCCCTCAC	870	
DB	563	AAGTGCATCATTAAGACTGTGGAGTTCGCCAACAGCTGCCCGGCTTACACCCCTCAC	722	
QY	871	ATCCCGACACAGATCAACCTTCCCTCAAGGCTGCTGCTGGACATCTGATCCTGCGGATC	930	
DB	723	ATCCCGACACAGATCAACCTTCCCTCAAGGCTGCTGCTGGACATCTGATCCTGCGGATC	782	
QY	931	TGCACGGGTACAGCCCGAGAGACACCATGACCTTCTCGGACGGGTGACCCCTGAAC	990	
DB	783	TGCACGGGTACAGCCCGAGAGACACCATGACCTTCTCGGACGGGTGACCCCTGAAC	842	
QY	991	CGGACCCAGATGCACAAAGCTGTGCTTGGCCCTTCCAGGACTTGGCTTTCGCTTCGC	1049	
DB	843	CGGACCCAGATGCACAAAGCTGTGCTTGGCCCTTCCAGGACTTGGCTTTCGCTTCGC	902	
QY	1050	CAACAGCTGTGCTGCCCTGGAGA	1072	
DB	903	CAACAGCTGTGCTGCCCTGGGA	925	

RESULT 7

AL551103/c

LOCUS

DEFINITION

993 bp

mRNA

linear

EST 16-FEB-2001

AL551103

AL551103 LTI\_NFL006\_PL2 Homo sapiens cDNA clone CS0D1066K12 3

prime, mRNA sequence.

ACCESSION AL551103

VERSION AL551103.1 GI:12888727

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 993)

AUTHORS Li, W.B., Gruber, C., Jesse, J. and Polayes, D.

TITLE Full-length cDNA libraries and normalization

COMMENT Unpublished (2001)

Contact: Genoscope

```

FEATURES
source
1. 993
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="GS0DI066YK12"
/clone_lib="LTI_NFL006_PL2"
/tissue_type="placenta"
/notes=Vector: pCMVSPORT 6; Site:1: NotI; 1st strand cDNA
was primed with a NotI-oligo(dT) primer. Five prime end
enriched double-stranded cDNA was digested with Not I and
cloned into the Not I and Eco RV sites of the pCMVSPORT 6
vector. Library was normalized. Library was constructed by
Life Technologies. Contact : Feng Liang Life Technologies,
a division of Invitrogen 9800 Medical Center Drive
Rockville, Maryland 20850, USA Fax : (1) 301 610 8371
Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com"
183 a 291 c 332 g 181 t 6 others
BASE COUNT

```

Query Match	38.7%	Score 807.2;	DB 9;	Length 993;
Best Local Similarity	99.1%;	Pred. No. 4.3e-105;		
Matches 853;	Conservative 0;	Mismatches 4;	Indels 4;	Gaps 4;
QY 1183	GCCTAAAGGCTCTACCTGCGAAGCGGAGCGCCACGCGCCCCACACATGTT-CCCCAGAT	1241		
Db				
861	GCCTAAATGCTCTACGTGCGGAAGCGGAGCGCCACGCGCCCCACACATGTTCCCCAAGAT	802		
QY 1242	GCTAATGAAGATTACTGACCTGCGAAGCATCAGCGCCAAGGGGCTCAGCGGGTGATCAC	1301		
Db				
801	GCTAATGAAGATTACTGACCTGCGAAGCATCAGCGCCAAGGGGCTCAGCGGGTGATCAC	742		
QY 1302	GCTGAAGATGGAGATCCCGGGCTCCATCGCGCTCTCATCCAGGAAATGTTGGAGAATC	1361		
Db				
741	GCTGAAGATGGAGATCCCGGGCTCCATCGCGCTCTCATCCAGGAAATGTTGGAGAATC	682		
QY 1362	AGAGGGCTTGGACACTCTAGCGGACAGCC- GGGGGGTGGGGGGCGGAGCGGGGTGGCC	1420		
Db				
681	AGAGGGCTTGGACACTCTAGCGGACAGCCGGGGGGGTGGGGGGCGGAGCGGGGTGGCC	622		
QY 1421	TGCCCCCCCCCGCAGGACGCTGTAGCCCCAGCCTCAGCCCCAGCTCCAACAGAAGCAGCC	1480		
Db				
621	TGCCCCCCCCCGCAGGACGCTGTAGCCCCAGCCTCAGCCCCAGCTCCAACAGAAGCAGCC	562		
QY 1481	CGGCACCCACTCCCGGTGACCGCCACGCCCACATGACACAGCCCTCGCCCTCCGCCCC	1540		
Db				
561	CGGCACCCACTCCCGGTGACCGCCACGCCCACATGACACAGCCCTCGCCCTCCGCCCC	502		
QY 1541	GCCTTTCTCTGCTTCTTACCGACCATGTGACCCCGCACCGCCCTGCCCTCC- ACCTGC	1599		
Db				
501	GCCTTTCTCTGCTTCTTACCGACCATGTGACCCCGCACCGCCCTGCCCTCC- ACCTGC	442		
QY 1600	CTCTCCGGGACGACTCTGGGACCTTCCCTTGGGGGACGGGGAGGAGGACGCGACTCC	1059		
Db				
441	CTCTCCGGGACGACTCTGGGACCTTCCCTTGGGGGACGGGGAGGAGGACGCGACTCC	382		
QY 1660	TTGGACAGAGGCGCTGGGGCCTCAGTGGACTGCTCTCCACAGCGCTGGGCTCAGCTCAG	1719		

RESULT 7						
AL551103/c						
LOCUS						
DEFINITION	AL551103	LTI_NFL006_P12	Homo sapiens	cDNA clone CS01066YK12_3	linear	EST 16-FEB-2001

```
|||||
Db 381 TTGGACAGAGGCGCTGGGCGCTCAGTGAGACTGCTGCTCCACAGCCTGGGCTGACGTCAG 322
QY 1720 AGGCCAGGCGCAGAACTAGTAGTGAGGCGCCCTGGTCTGGGCTCAGGATGGGTCCTGGGG 1779
Db 321 AGGCCAGGCGCAGAACTAGTAGTGAGGCGCCCTGGTCTGGGCTCAGGATGGGTCCTGGGG 262
QY 1780 GCTCTGCTGTCATCAGACACACCCCTCTGCCAGCTCAGCAGCATCTTCATCAGCAGCAAC 1839
Db 261 GCTCTGCTGTCATCAGACACACCCCTCTGCCAGCTCAGCAGCATCTTCATCAGCAGCAAC 202
QY 1840 GCCAGGACTTGGCTCCGCCCTCCTCAGAACTCACAAGCCATGCTCCCGAGCTGGGGAAC 1899
Db 201 GCCAGGACTTGGCTCCGCCCTCCTCAGAACTCACAAGCCATGCTCCCGAGCTGGGGAAC 142
QY 1900 CTCACCTCCCGCTGCTGCTGGTGTGACAGAGGGGGTGGGAGAGGGGGGGGGTTC 1959
Db 141 CTCACCTCCCGCTGCTGCTGGTGTGACAGAGGGGGTGGGAGAGGGGGGGGGTTC 82
QY 1960 CCTGTACATACCTCCCATACCAACCCAGGTATTAATTCCTGCTGGTGTGTTTTT-A 2018
Db 81 CCTGTACATACCTCCCATACCAACCCAGGTATTAATTCCTGCTGGTGTGTTTTTNA 22
QY 2019 TTTTAAATTTTGTGTTTGTAT 2039
Db 21 TTTTAAATTTTGTGTTTGTAT 1

RESULT 8
BQ706025
LOCUS BQ706025 885 bp mRNA linear EST 16-JUL-2002
DEFINITION AGENCOURT_8351580 NIH_MGC_113 Homo sapiens cDNA clone IMAGE:6282394
5', mRNA sequence.
ACCESSION BQ706025
VERSION BQ706025.1 GI:21844924
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS NIH-MGC http://mgi.nhl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgabbs-remail.nhl.nih.gov
Tissue Procurement: Dr. Mark Watson
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCM2476 row: 1 column: 11
High quality sequence stop: 599.
Location/Qualifiers
1..885
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:6282394"
/lab_host="NIH_MGC_113"
/note="Organ: spleen; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGGAG(G). Library constructed by Ling Hong in the
Laboratory of Gerald M. Rubin (University of California,
Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and
Superscript II RT (Life Technologies). Note: this is a
NIH_MGC Library."
BASE COUNT 157 a 337 c 234 g 154 t 3 others
ORIGIN
```

```
Query Match 35.08; Score 729.6; DB 14; Length 885;
Best Local Similarity 91.08; Pred. No. 4.2e-94;
Matches 785; Conservative 0; Mismatches 77; Indels 1; Gaps 1;

QY 678 GAAGGTGCGCAAGCGCACAGGAAACCTTCCCTGCCCTCTGCCAGCTGGGCAAAATACAC 737
Db 1 GAAGGTGCGCAAGCGCACAGGAAACCTTCCCTGCCCTCTGCCAGCTGGGCAAAATACAC 60
QY 738 TAGGAACAACAGCTCAGAACACGCTGTCTCTCTGTGACATTTGACCTCTGGGCAAAATACAC 797
Db 61 TAGGAACAACAGCTCAGAACACGCTGTCTCTCTGTGACATTTGACCTCTGGGCAAAATACAC 120
QY 798 TGAACCTCTCACCAAGTGCATCATTAAGACTGTGGAGTTGCGCAAGAGAGCTGCCGGCTT 857
Db 121 TGAACCTCTCACCAAGTGCATCATTAAGACTGTGGAGTTGCGCAAGAGAGCTGCCGGCTT 180
QY 858 CACCACCTCTCACCATCGCGACAGATCACCCTCTCTCAAGGCTGCTGCTGGACATCTCT 917
Db 181 CACCACCTCTCACCATCGCGACAGATCACCCTCTCTCAAGGCTGCTGCTGGACATCTCT 240
QY 918 GATCCTCGGATCTGCACGCGGTACACGCCGAGCAGACACCATGACCTTCTCGGACGG 977
Db 241 GATCCTCGGATCTGCACGCGGTACACGCCGAGCAGACACCATGACCTTCTCGGACGG 300
QY 978 GCTGACCTGAAACCGGACCCAGATGCACACGCTGGCTTCCGCCCTCTACCGACCTGGT 1037
Db 301 GCTGACCTGAAACCGGACCCAGATGCACACGCTGGCTTCCGCCCTCTACCGACCTGGT 360
QY 1038 CTTTGGCTTCGCCAACCCAGCTGCTGCCCTGGAGATGATGCGGAGAGCGGGCTGCT 1097
Db 361 CTTTGGCTTCGCCAACCCAGCTGCTGCCCTGGAGATGATGCGGAGAGCGGGCTGCT 420
QY 1098 CAGCGCCATCTGCTCATCTGCGGAGACCGCGCAGGACCTGAGGAGCGCCGCGGTGGA 1157
Db 421 CAGCGCCATCTGCTCATCTGCGGAGACCGCGCAGGACCTGAGGAGCGCCGCGGTGGA 480
QY 1158 CATGCTCAGAGAGCGCTGCTGGAGCGCTAAAGGTCTACGTGGGAAGCGGAGCCCGAG 1217
Db 481 CATGCTCAGAGAGCGCTGCTGGAGCGCTAAAGGTCTACGTGGGAAGCGGAGCCCGAG 540
QY 1218 CCGCCCGACATGTTCCCAAGATGCTAAAGATTAATGACCTGCGCAAGCATCAGCGC 1277
Db 541 CCGCCCGACATGTTCCCAAGATGCTAAAGATTAATGACCTGCGCAAGCATCAGCGC 600
QY 1278 CAAGGGGCTGAGCGGTGATCAGCTGAAGATGAGATCCCGGCTCCATGCCGCTCT 1337
Db 601 CAAGGGGCTGAGCGGTGATCAGCTGAAGATGAGATCCCGGCTCCATGCCGCTCT 660
QY 1338 CATCCAGGAATGTT-GGAGAACTCAGAGGCGCTGGACACTCTGAGCGGACAGCGGGGG 1396
Db 661 CATCCAGGAATGTTGGGAGAACTCAGAGGCGCTGGACACTCTGAGCGGACAGCGGGGG 720
QY 1397 GTGGGGGGGGGAGGGGGTGGCTGCCCGCCCGCAGGAGCTGTAGTCCCGACCTCA 1456
Db 721 GTGGGGGGGGGNNCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 780
QY 1457 GCCCGAGCTCCAAAGAGAGCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 1516
Db 781 GCCCGCTCCNCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 840
QY 1517 GACACAGCGCTCCCGCTCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 1539
Db 841 GCCCGTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 863

RESULT 9
BC030234
LOCUS BC030234
DEFINITION Homo sapiens, retinoic acid receptor, beta, clone IMAGE:5229160,
mRNA.
ACCESSION BC030234
VERSION BC030234.1 GI:20988815
KEYWORDS HTC.
```

SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS 1 (bases 1 to 3141)  
TITLE Strausberg, R.  
JOURNAL Direct Submission  
Submitted (07-MAY-2002) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
COMMENT Contact: MGC help desk  
Email: [cgapbs-remail.nih.gov](mailto:cgapbs-remail.nih.gov)  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Prepared by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: National Institutes of Health Intramural  
Sequencing Center (NISC),  
Gaithersburg, Maryland;  
Web site: <http://www.nisc.nih.gov/>  
Contact: [nisc\\_mgc@nhgri.nih.gov](mailto:nisc_mgc@nhgri.nih.gov)  
Akhter, N., Ayele, K., Beckstrom-Sternberg, S.M., Benjamin, B.,  
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,  
Dietrich, N.L., Granite, S., Guan, X., Gupta, J., Haghghi, P.,  
Hansen, N., Ho, S.-L., Karlins, E., Laric, P., Legaspi, R., Maduro, Q.L.,  
Masello, C., Masker, B., Mastrian, S.D., McCloskey, J.C., McDowell, J.,  
Pearson, R., Stanripop, S., Thomas, P.J., Touchman, J.W., Tsurgon, C.,  
Vogt, J.L., Walker, M.A., Wetherby, K.D., Wiggins, L., Young, A.,  
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
Series: IRAC Plate: 49 Row: p Column: 20  
This clone was selected for full length sequencing because it  
passed the following selection criteria: matched mRNA gi: 14916493  
This clone has the following problem: frame shifted.

FEATURES  
source  
1..3141  
/organism="Homo sapiens"  
/db\_xref="LocusID:5915"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5229160"  
/tissue\_type="Pancreas, Spleen, adult pooled"  
/clone\_lib="NIH\_MGC\_120"  
/lab\_host="DH10B"  
/note="vector: pCMV-SPORT6"  
BASE COUNT 986 a 663 c 633 g 859 t  
ORIGIN

Query Match 32.5%; Score 678.8; DB 11; Length 3141;  
Best Local Similarity 76.1%; Pred. No. 2.7e-87;  
Matches 836; Conservative 0; Mismatches 262; Indels 0; Gaps 0;  
Qy 286 CACTCATTGAGACCCAGCAGCAGTCTGAGAGATAGTCCAGAGCCCTCCCTCGCA 345  
Db 543 CACTCATTGAGACCCAGCAGCAGTCTGAGAGATAGTCCAGAGCCCTCCCTCGCA 602  
Qy 346 CCCCCTTACCCCGCATCAAGCCTTGTCTGTGAGGACCAAGTCCCTCAGGCTAC 405  
Db 603 CTCTCCCTCCCTCGAGTGTACAAACCTCTGTCTGTCTCCAGGACCAATCATCAGGGTAC 662  
Qy 406 CACTATGGGGTTCAGCGCTGTGAGGGTGTCAAGGGCTTCTTCCGCGCAGCATCCAGAG 465  
Db 663 CACTATGGGGTTCAGCGCTGTGAGGGTGTCAAGGGCTTCTTCCGCGCAGAGTATTCAGAG 722  
Qy 466 AACATGGTGTACAGTGTACCCGGGACCAAGTGTCTCATCAACCAAGTGTACCCGGGAC 525  
Db 723 AATATGATTTTACATGTGTACCCGGGACCAAGTGTCTTATTAATAAAGTCAACAGGAAT 782  
Qy 526 CCCTCGAGTACTCCGAGTCCAGAGTGTTCAGTGGGCGCATGTCCAGGAGTCTGTG 585  
Db 783 CGATGCCAATACTGTGCACTCCAGAGTGTTCAGAGTGGGAATGTCCAAAGAAATCTGTC 842

QY 586 AGAAGCAGCCGAACAGAGAGAGAGTGTCCCAAGCCCGAGTGTCTCTGAGAGCTAC 645  
Db 843 AGAATGACAGGACACAGAGAGAGAGTGTCCCAAGCCCGAGTGTCTCTGAGAGCTAT 902  
QY 646 ACCTGACGCGCGGAGGTGGGGAGCTCATTTGAGAAGTGGGCAAGGCGCACAGGAAACC 705  
Db 903 GAAATGACAGCTGAGTTGGAGCATCTCACAGAGAAGATCCGAAAAGCTCACCAGAACT 962  
QY 706 TTCCCTGCCCTTGCAGCTGGGCAATACATACAGAACACAGCTCAGAACACAGTGTG 765  
Db 963 TTCCCTTCACTTGCAGCTGGGTAATACACACAGAAATCCAGTGTGACCATCGAGTC 1022  
QY 766 TCTCTGACATTTGACCTCTGAGCAAGTTCAGTGAACCTCTCCACCAAGTGCATCATTAAG 825  
Db 1023 CGACTGAGCTGGGCTCTGGGCAATTCAGTGAACCTGGGCAAGTGCATCATTAAG 1082  
QY 826 ACTGTGAGTTCGCCAAGCAGCTGCCGGCTTCACCAACCTCACCATCGCGGACGAGTC 885  
Db 1083 ATCGTGGAGTTTGTAAACGCTGTGCTGTGTTTCACTGGCTTGACCATCGCAGCAAAAT 1142  
QY 886 ACCCTCTCAAGCTGCCCTGGAGATCTCTGCGGATCTGCGCGGTGACGCGGTACAG 945  
Db 1143 ACCCTGCTGAAGCGCGCTGCCCTGGAGATCTCTGAGTGTGTTAGAATTTGACCAAGTATACC 1202  
QY 946 CCGAGCAGGACACCATGACCTTCTCGGAGCGGTGACCTGAAACCGGACCCAGATGCAC 1005  
Db 1203 CCAGAACAGACACCATGACTTCTCAGAGCGGCTTACCTTAATCGAACTCAGATGCAC 1262  
QY 1006 AAGCTGGCTTGGGCCCCCTCAGCAGCTGCTTGGCTTTCGCCCAACAGCTGCTGCC 1065  
Db 1263 AATGCTGGATTTGGTCTCTGACTGACCTGTGTTTCACTTGGCAACAGCTCCTGCC 1322  
QY 1066 CTGGAGATGATGATCGGAGCGGCTGCTCAGCGGCTGCTGCTCATCTCGGAGAC 1125  
Db 1323 TTGGAATGGATGACAGAAACAGCGCTTCTCAGTGGCATCTGCTTAATCTGGAGAC 1382  
QY 1126 CGCCAGGACCTGAGCAGCGGCGGTGGACATGCTGCGAGGCGCTGCTGGAGGCG 1185  
Db 1383 CGCCAGGACCTTGGAGAACCCAGCAAAAGTAGATAAGCTACAAGAACCATTTGCTGGA 1442  
QY 1186 CTAAAGGCTAGTGGGAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1245  
Db 1443 CTAAATTTTATATCAGAAAAAGACGACCGGCGGCGGCGGCGGCGGCGGCGGCGG 1502  
QY 1246 ATGAAGATTACTGACCTGGAAGCATCAGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1305  
Db 1503 ATGAATTCACAGATCTCCGTAGCATGCTGCTAAAGGTGACAGCGGTGTAATACCTG 1562  
QY 1306 AAGATGGAGATCCCGGCTTCCATGCGCTCTCATCCAGAAATGTTGGAGAACTCAGAG 1365  
Db 1563 AAAATGGAATTCCTGGATCAATGCCACCTCTCATTCAGAAATGCTGGAGAAATCTGAA 1622  
QY 1366 GCCTGGACACTCTGAGC 1383  
Db 1623 GGACATGAACCTTGACC 1640  
RESULT 10  
BI827961  
LOCUS 603073894F1 NIH\_MGC\_119 Homo sapiens cDNA clone IMAGE:5165734.5',  
DEFINITION mRNA sequence.  
ACCESSION BI827961  
VERSION BI827961.1 GI:15939511  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 702)  
AUTHORS NIH-MGC <http://mgc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)



JOURNAL COMMENT

Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Life Technologies, Inc.  
cDNA Library Preparation: Life Technologies, Inc.  
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone Distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLAM11411 row: 1 column: 23  
High quality sequence stop: 622.  
Location/Qualifiers  
1. 702  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_lib="NIH\_MGC\_119"  
/tissue\_type="medulla"  
/lab\_host="DH10B"  
/note="Organ: brain; Vector: pCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA source normal medulla from anonymous male age 27. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.3 kb, insert size range 0.9-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 013. Note: this is a NIH\_MGC Library."  
BASE COUNT 145 a 237 c 212 g 108 t  
ORIGIN

Query Match 31.6%; Score 659.8; DB 13; Length 702;  
Best Local Similarity 98.0%; Pred. No. 3.6e-84;  
Matches 689; Conservative 0; Mismatches 12; Indels 2; Gaps 2;  
QY 794 TCAGTGAACCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCG 853  
DB 1 TCAGTGAACCTCCACCAAGTGCATCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCG 60  
QY 854 GCTTACCACCTCCACCATCGCGACCATGATCACCCTCCCTCAAGCTGCCTGCTGACCA 913  
DB 61 GCTTACCACCTCCACCATCGCGACCATGATCACCCTCCCTCAAGCTGCCTGCTGACCA 120  
QY 914 TCCTGATCTCTCGGATCTGCACGCGGTACACGCGGAGCAGGACCATGACCTTCCTCGG 973  
DB 121 TCCTGATCTCTCGGATCTGCACGCGGTACACGCGGAGCAGGACCATGACCTTCCTCGG 180  
QY 974 ACGGCTGACCTTGAACCGGACCCAGATGATGACACGCTGCTGCGCCCTCCACCGACC 1033  
DB 181 ACGGCTGACCTTGAACCGGACCCAGATGATGACACGCTGCTGCGCCCTCCACCGACC 240  
QY 1034 TGGTCTTTCCTTCGCGCAACCACTGCTGCTGCGGAGATGGATGATGCGGAGACGGGG 1093  
DB 241 TGGTCTTTCCTTCGCGCAACCACTGCTGCTGCGGAGATGGATGATGCGGAGACGGGG 300  
QY 1094 TGCTACGCGCATCTGCTCATCTGCGGAGACCGCAGGACCTGGAGACCGCGGACCGGG 1153  
DB 301 TGCTACGCGCATCTGCTCATCTGCGGAGACCGCAGGACCTGGAGACCGCGGACCGGG 360  
QY 1154 TGGACATGCTGACGAGGCGCTGCTGAGGCGCTAAAGTCTACGTGCGGAAGCGGAGGC 1213  
DB 361 TGGACATGCTGACGAGGCGCTGCTGAGGCGCTAAAGTCTACGTGCGGAAGCGGAGGC 420  
QY 1214 CCAGCGCCCCACATGTTCCCCAAGATGCTAATGAAGTACTGACCTCGGAAGCATCA 1273  
DB 421 CCAGCGCCCCACATGTTCCCCAAGATGCTAATGAAGTACTGACCTCGGAAGCATCA 480  
QY 1274 GCGCCAAAGGGGCTGACGGGTGATCAGCTGAAGATGGAGATCCCGGGTCCATGCCCG 1333  
DB 481 GCGCCAAAGGGGCTGACGGGTGATCAGCTGAAGATGGAGATCCCGGGTCCATGCCCG 540  
QY 1334 CTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCCCTGGACACTCTCAGCGGACGCGG 1393

DB 541 CTCTCATCCAGGAAATGTTGGAGAACTCAGAGGGCTGGAACTCTGAGCGGACGCGG 600  
QY 1394 GGGGTGGGGGGCGGACGGGTGGCTCCCGCCCGCCCGCCAGCAGCT-CTAGCCCCCAGC 1452  
DB 601 GGGGTGGGGGGCGGACGGGTGGCTGGCTGGCCCGCCCGCCAGCAGCTAGTAGCCCCCAGG 660  
QY 1453 CTACGCCCGCAGCTCCCAACAGAGCAGCGCGGCCACCCACTCTCC 1495  
DB 661 CTCAGCCCAAGATCC-ACAGAAGCAGCGCGGCAACACCTACCC 702

RESULT 11  
BQ044936/c  
LOCUS  
DEFINITION  
BQ044936  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

654 bp mRNA linear EST 28-MAR-2002  
UI-H-EU0-azo-e-05-0-UI-s1 NCI\_CGAP\_Carl Homo sapiens cDNA clone  
IMAGE:5851372 3', mRNA sequence.  
BQ044936  
BQ044936.1 GI:19795852  
EST.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 654)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: Dr. Jose Mercuende  
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
DNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa  
Clone Sequencing by: Dr. M. Bento Soares, University of Iowa  
Clone Distribution: Clone distribution information can be found  
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov  
The following repetitive elements were found in this cDNA  
sequence: 1-73, >AT-rich#low\_complexity (matched complement)  
Seq primer: M13 FORWARD  
POLYA-Yes.

Location/Qualifiers  
1. 654  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5851372"  
/clone\_lib="NCI\_CGAP\_Carl"  
/tissue\_type="Osteoarthritic Cartilage"  
/dev\_stage="adult"  
/lab\_host="DH10B (Life Technologies)"  
/note="Organ: Knee; Vector: pT7T3-Pac (Pharmacia) with a modified polylinker; Site\_1: EcoR I; Site\_2: Not I; NCI\_CGAP\_Carl is a cDNA library containing the following tissue(s): Osteoarthritic Cartilage The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dfr)18 tail. The sequence tag for this library is TGATCAGCGCT.  
TAG LIB=UI-H-EU0  
TAG TISSUE=osteothritic cartilage  
TAG\_SEQ=TCATCAGCGCT"

BASE COUNT 135 a 161 c 227 g 131 t  
ORIGIN  
Query Match 31.4%; Score 654; DB 14; Length 654;  
Best Local Similarity 100.0%; Pred. No. 2.5e-83;  
Matches 654; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1432 CCAGCAGCTGTAGCCAGCCTCAGCCAGCTCAACAGAGCAGCCCGCCACCCAC 1491
|
|
|
Db 654 CCAGCAGCTGTAGCCAGCCTCAGCCAGCTCAACAGAGCAGCCCGCCACCCAC 595
|
|
|
QY 1492 TCCCGGTGACCGCCACCGCCACATGGACACAGCCTCGCCCTCCCGCGGCTTTCTCT 1551
|
|
|
Db 594 TCCCGGTGACCGCCACCGCCACATGGACACAGCCTCGCCCTCCCGCGGCTTTCTCT 535
|
|
|
QY 1552 GCCTTTCTACCGACCATGTGACCCCGCAGCCTGCCCCACCTGCCCCGCGGAG 1611
|
|
|
Db 534 GCCTTTCTACCGACCATGTGACCCCGCAGCCTGCCCCACCTGCCCCGCGGAG 475
|
|
|
QY 1612 TACTGGGACCTTCCCTGGGGACGGGAGGAGGAGGAGCAGCAGTCTCTTGACAGAGC 1671
|
|
|
Db 474 TACTGGGACCTTCCCTGGGGACGGGAGGAGGAGGAGCAGCAGTCTCTTGACAGAGC 415
|
|
|
QY 1672 CTGGGCCCTCAGTGAGCTGCTGCTCCACAGCCTGGGCTGACGTGACAGGCCAGGCCA 1731
|
|
|
Db 414 CTGGGCCCTCAGTGAGCTGCTGCTCCACAGCCTGGGCTGACGTGACAGGCCAGGCCA 355
|
|
|
QY 1732 GGAAGTGTAGTGAGGCCCTGTGCTCTGGGTCTCAGGATGGTCTCGGGGCTCTGTCA 1791
|
|
|
Db 354 GGAAGTGTAGTGAGGCCCTGTGCTCTGGGTCTCAGGATGGTCTCGGGGCTCTGTCA 295
|
|
|
QY 1792 TCAAGACACCCCTCTGCCAGCTCACCACATCTTCTCACCAGCAAGCCAGGACTTGG 1851
|
|
|
Db 294 TCAAGACACCCCTCTGCCAGCTCACCACATCTTCTCACCAGCAAGCCAGGACTTGG 235
|
|
|
QY 1852 CTCCCCATCTCAGAACTCAGACCCATGTCTCCAGCTGGGAACTCAACCTCCCC 1911
|
|
|
Db 234 CTCCCCATCTCAGAACTCAGACCCATGTCTCCAGCTGGGAACTCAACCTCCCC 175
|
|
|
QY 1912 CTGCTCGTGTGTGACAGAGGGGTGGACAGGGGGGGGGTTCCTCTGTACATAC 1971
|
|
|
Db 174 CTGCTCGTGTGTGACAGAGGGGTGGACAGGGGGGGGGTTCCTCTGTACATAC 115
|
|
|
QY 1972 CTGCCATACCAACCCAGGTATTAATCTCGTGTGTGTGTATTTATTTATTTT 2031
|
|
|
Db 114 CTGCCATACCAACCCAGGTATTAATCTCGTGTGTGTGTATTTATTTATTTT 55
|
|
|
QY 2032 GTTTTGTATTTTATTAAGAAATTTTCAATTTAAGCACAAAAA 2085
|
|
|
Db 54 GTTTTGTATTTTATTAAGAAATTTTCAATTTAAGCACAAAAA 1

RESULT 12
BM927280
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1091)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-re@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM12827 row: n column: 08
High quality sequence stop: 621.
```

FEATURES  
source

Location/Qualifiers

```
1..1091
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5768263"
/clone_lib="NIH_MGC_121"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: EcoRV (destroyed); RNA source anonymous pool of 3
fetal brains, female age 20 weeks, female age 24 weeks,
and male age 26 weeks. Library is oligo-dr primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.7 kb, insert size range
0.7-3.5 kb. Library is normalized and enriched for
full-length clones and was constructed by C. Gruber
(invitrogen). Research Genetics tracking code 017. Note:
this is a NIH_MGC Library."
BASE COUNT 278 a 339 c 286 g 188 t
```

```
Query Match 31.1%; Score 648.4; DB 14; Length 1091;
Best Local Similarity 98.3%; Pred. No. 1.1e-82;
Matches 655; Conservative 0; Mismatches 11; Indels 0; Gaps 0;
```

```
QY 1420 CTGCCCCCCCCCAGGAGCTGTAGCCAGCCTCAGCCCCAGCTCCACAGAGCAGC 1479
|
|
|
Db 1 CTGCCCCCCCCCAGGAGCTGTAGCCAGCCTCAGCCCCAGCTCCACAGAGCAGC 60
|
|
|
QY 1480 CGGGCCACCCACTCCCGGTGACCCGCCACGACATGGACAGCCCTCGCCCTCGGCC 1539
|
|
|
Db 61 CGGGCCACCCACTCCCGGTGACCCGCCACGACATGGACAGCCCTCGCCCTCGGCC 120
|
|
|
QY 1540 CGGCTTTCTCTGCTTTCTACCGACCATGTACCCGCCAGCCCTGCCCCACCTGC 1599
|
|
|
Db 121 CGGCTTTCTCTGCTTTCTACCGACCATGTACCCGCCAGCCCTGCCCCACCTGC 180
|
|
|
QY 1600 CCTCCCGGCGAGTCTGCGGACCTTCCCTGGGGAGCGGGAGGAGGAGCGAGCTCC 1659
|
|
|
Db 181 CCTCCCGGCGAGTCTGCGGACCTTCCCTGGGGAGCGGGAGGAGGAGCGAGCTCC 240
|
|
|
QY 1660 TTGGACAGAGGCTGGGCGCTCAGTGGACTGCTGCTCCACAGCCCTGGGCTGACGTCA 1719
|
|
|
Db 241 TTGGACAGAGGCTGGGCGCTCAGTGGACTGCTGCTCCACAGCCCTGGGCTGACGTCA 300
|
|
|
QY 1720 AGGCCGAGGCGAGGAACCTGAGTGAGGCCCTCGTCTGGGTCTCAGATGGGTCTGGGG 1779
|
|
|
Db 301 AGGCCGAGGCGAGGAACCTGAGTGAGGCCCTCGTCTGGGTCTCAGATGGGTCTGGGG 360
|
|
|
QY 1780 GCCTGCTGTTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAAC 1839
|
|
|
Db 361 GCCTGCTGTTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAAC 420
|
|
|
QY 1840 GCCAGGACTTGGCTCCCCCATCTCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAAC 1899
|
|
|
Db 421 GCCAGGACTTGGCTCCCCCATCTCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAAC 480
|
|
|
QY 1900 CTCACCTCCCGGCTCGGTTGGTGACAGAGGGGTGGGAGGAGGCGGGGGTTC 1959
|
|
|
Db 481 CTCACCTCCCGGCTCGGTTGGTGACAGAGGGGTGGGAGGAGGCGGGGGTTC 540
|
|
|
QY 1960 CCCTGTACATACCTGCCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTTTAT 2019
|
|
|
Db 541 CCCTGTACATACCTGCCATACCAACCCAGGTATTAATTCGCTGGTGTGTTTTTAT 600
|
|
|
QY 2020 TTTATTTTGTGTTGATTTTATTAAGAAATTTTCAATTTAAGCACAAAAA 2079
|
|
|
Db 601 TTTATTTTGTGTTGATTTTATTAAGAAATTTTCAATTTAAGCACAAAAA 660
|
|
|
QY 2080 AAAAAA 2085
|
|
|
Db 661 AAGGAA 666
```



```
RESULT 13
AI806984/c
LOCUS      AI806984      703 bp      mRNA      linear      EST 18-DEC-1999
DEFINITION w24908.x1 Soares_NFL_T_GBC_S1 Homo sapiens cDNA clone
IMAGE:2356574 3' similar to gb:M73779 RETINOIC ACID RECEPTOR
ALPHA-1 (HUMAN) ; mRNA sequence.
ACCESSION  AI806984
VERSION     AI806984.1 GI:5393550
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 703)
            NCI-CCAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            Insert length: 794 Std Error: 0.00.
            Seq primer: -400p from Gibco
            High quality sequence stop: 469.
FEATURES   Location/Qualifiers
            1..703
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:2356574"
            /clone_lib="Soares_NFL_T_GBC_S1"
            /lab_host="DH10B"
            /note="Organ: pooled; Vector: pTT3D-Pac (Pharmacia) with
            a modified polylinker; Site_1: Not I; Site_2: Eco RI;
            Equal amounts of plasmid DNA from three normalized
            libraries (fetal lung NBHL19W, testis NHT, and B-cell
            NCI CGAP GCBI) were mixed, and ss circles were made in
            vitro. Following HAP purification, this DNA was used as
            tracer in a subtractive hybridization reaction. The driver
            was PCR-amplified cDNAs from pools of 5,000 clones made
            from the same 3 libraries. The pools consisted of
            I.M.A.G.E. clones 297480-302087, 682632-687239,
            726408-728711, and 729096-731399. Subtraction by Bento
            Soares and M. Fatima Bonaldo."
BASE COUNT 144 a 184 c 240 g 133 t 2 others
ORIGIN
Query Match 30.9%; Score 645.4; DB 9; Length 703;
Best Local Similarity 96.5%; Pred. No. 3.8e-82;
Matches 680; Conservative 0; Mismatches 23; Indels 2; Gaps 2;
QY 1372 GACACTGTAGCGGACAGCGGGGGTGGGGGGGGGACGCGGGGTGGCGCTGCCCGCCCGC 1431
Dbb 703 GACATTCTGTAGCGAAGCAGCAGAGAGT-GAGAGCGAGAGCGTAGTGG-CTGGGCGCCCGC 646
QY 1432 CCAGGACGCTGTAGCCCGCAGCTCAGCCCGGCTCCACAGAGCAGCGCCCGCCACCCAC 1491
Dbb 645 CAGGGCAGCTGTAGCCCGCAGCTCAGCCCGGCTCCACAGAGCAGCGCCCGCCACCCAC 586
QY 1492 TCCCGGTGACCGCCCGCAGCGGACATGGACACAGCGCTCGCCCTCCGCGCGCGCTTTCTCT 1551
Dbb 585 TCNCGGTGACCGCCCGCAGCGGACATGGACACAGCGCTCGCCCTCCGCGCGCGCTTTCTCT 526
QY 1552 GCCTTTTACCGACCATGTGACCCCGGACCGAGCGCGCTCGCCCGCCCGCCTCGCCCGGCGAG 1611
Dbb 525 GCCTTTTACCGACCATGTGACCCCGGACCGAGCGCGCTCGCCCGCCCGCCTCGCCCGGCGAG 466
QY 1612 TACTGGGAGCTTCCCTGCGGGGAGCGGGAGGAGGAGCGGAGCGGAGCGGAGCGGAGCGGAG 1671
Dbb 465 TACTGGGAGCGCTTCCCTGCGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 406
QY 1672 CTGGGCGCTCAGTGGAGCTGCTGCTCCACAGCGCTGGGCTGACCTCAGAGCGCCGAGGCCA 1731
|||||
```

```
Dbb 405 CTGGGCGCTCAGTGGAGCTGCCTGCTCCACAGCGCTGGGCTGAGCTCAGAGCGCGAGGCCA 346
QY 1732 GGAACCTGAGTGGAGCGCCCTGGTCTGGGGTCTCAGGATGGGTCTCTGGGGCGCTCGTGTTC 1791
|||||
Dbb 345 GGAACCTGAGTGGAGCGCCCTGGTCTGGGGTCTCAGGATGGGTCTCTGGGGCGCTCGTGTTC 286
QY 1792 TCAAGACACCCCTCTGCCAGCTCCACCATCTTATCATCACCAGCAACGCCAGGACTTGG 1851
|||||
Dbb 285 TCAAGACACCCCTCTGCCAGCTCCACCATCTTATCATCACCAGCAACGCCAGGACTTGG 226
QY 1852 CTCGCCCATCTCTCAGAACTCACAAGCCATTTGCTCCCGAGCTGGGGAACTCAACCTCCCC 1911
|||||
Dbb 225 CTCGCCCATCTCTCAGAACTCACAAGCCATTTGCTCCCGAGCTGGGGAACTCAACCTCCCC 166
QY 1912 COTGCTCGTGGTGGTGCAGAGGGGGTGGGACAGGGGGGGGGGGTTCCTCCCTGTGTACAT 1971
|||||
Dbb 165 COTGCTCGTGGTGGTGCAGAGGGGGTGGGACAGGGGGGGGGGGTTCCTCCCTGTGTACAT 106
QY 1972 COTGCTCATACCAACCCAGGATTAATTCCTCGCTGTTTTGTTTTTATTTTATTTTATTTT 2031
|||||
Dbb 105 COTGCTCATACCAACCCAGGATTAATTCCTCGCTGTTTTGTTTTTATTTTATTTTATTTT 46
QY 2032 GTTTTGATTTTTTTTAAAGAAATTTTCATTTTAAAGCAACAAAAA 2076
|||||
Dbb 45 GTTCTGATTTTTTTTAAAGAAATTTTCATTTTAAAGCAACATTTAAA 1
RESULT 14
BG028346
LOCUS      BG028346      1164 bp      mRNA      linear      EST 24-JAN-2001
DEFINITION 602295596F1 NIH_MGC_86 Homo sapiens cDNA clone IMAGE:4390510 5',
            mRNA sequence.
ACCESSION  BG028346
VERSION     BG028346.1 GI:12417440
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 1164)
            NIH-MGC http://mgc.nci.nih.gov/.
            National Institutes of Health, Mammalian Gene Collection (MGC)
            Unpublished (1999)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-re@mail.nih.gov
            Tissue Procurement: ATCC
            CDNA Library Preparation: Life Technologies, Inc.
            CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10079 row: 0 column: 23
            High quality sequence stop: 677.
FEATURES   Location/Qualifiers
            1..1164
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="IMAGE:4390510"
            /clone_lib="NIH_MGC_86"
            /tissue_type="osteosarcoma, cell line"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: bone; Vector: pCMV-Sport6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally; oligo-dT primed.
            Average insert size 1.533 kb. Library enriched for
            full-length clones and constructed by Life Technologies.
            Note: this is a NIH_MGC Library."
BASE COUNT 285 a 376 c 339 g 164 t
ORIGIN
Query Match 30.6%; Score 639; DB 12; Length 1164;
Best Local Similarity 97.4%; Pred. No. 2.2e-81;
Matches 681; Conservative 0; Mismatches 15; Indels 3; Gaps 3;
```

QY 1330 CCGCCTCTCATCCAGAAATGTTGGAACTCAGAGGCGCTGGACACTCTGAGCGGACAG 1389  
|||||  
Db 1 CCGCCTCTCATCCAGAAATGTTGGAACTCAGAGGCGCTGGACACTCTGAGCGGACAG 60  
|||||  
QY 1390 CCGGGGGTGGGGGGGGGAGCGGGGTGGCTGCCCGCCCGCCAGGACGCTGTAGCCCC 1449  
|||||  
Db 61 CCGGGGGTGGGGGGGGGAGCGGGGTGGCTGCCCGCCCGCCAGGACGCTGTAGCCCC 120  
|||||  
QY 1450 AGCCTCAGCCCCAGCTCCCAACAGACAGCCCGGCCACCCACCTCCCGTACCGCCCGCAG 1509  
|||||  
Db 121 AGCCTCAGCCCCAGCTCCCAACAGACAGCCCGGCCACCCACCTCCCGTACCGCCCGCAG 180  
|||||  
QY 1510 CCACATGAGACAGCCCTCGCCCTCGCCCGCCGCTTTCTCTGCTTCTTACCGACCATG 1569  
|||||  
Db 181 CCACATGAGACAGCCCTCGCCCTCGCCCGCCGCTTTCTCTGCTTCTTACCGACCATG 240  
|||||  
QY 1570 TGACCCCGACAGCCCTCGCCCGCCAGCTGCCCTCCCGGGGAGTACGAGGACCTTCCCTG 1629  
|||||  
Db 241 TGACCCCGACAGCCCTCGCCCGCCAGCTGCCCTCCCGGGGAGTACGAGGACCTTCCCTG 300  
|||||  
QY 1630 GGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1689  
|||||  
Db 301 GGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 360  
|||||  
QY 1690 GCCTGCTCCACAGCCTGGGCTGACGCTCAGAGGCCGAGGCCAGGAACTGAGTGAGGCCCC 1749  
|||||  
Db 361 GCCTGCTCCACAGCCTGGGCTGACGCTCAGAGGCCGAGGCCAGGAACTGAGTGAGGCCCC 420  
|||||  
QY 1750 TGGTCTGGGTCTCAGGATGGGTCTCTGGGGGCTCGTGTTCATCAAGACACCCCTCTGCC 1809  
|||||  
Db 421 TGGTCTGGGTCTCAGGATGGGTCTCTGGGGGCTCGTGTTCATCAAGACACCCCTCTGCC 480  
|||||  
QY 1810 CAGCTCACCACATCTTCATCACCAGCAAGAGCGGAGGAGTGGTCCCGCATCTCAGAAC 1869  
|||||  
Db 481 CAGCTCACCACATCTTCATCACCAGCAAGAGCGGAGGAGTGGTCCCGCATCTCAGAAC 540  
|||||  
QY 1870 TCACAGGACCATCTGCTCCCGAGTGGGAACTCAACCTCCCGCTGCC-TCGGTGGTGA 1928  
|||||  
Db 541 TCACAGGACCATCTGCTCCCGAGTGGGAACTCAACCTCCCGCTGCC-TCGGTGGTGA 600  
|||||  
QY 1929 CAGAGGGGTGGGACAGGGG-CGGGGGTTCCCGCTGTATACCATCCCTG-CCATACCAAC 1986  
|||||  
Db 601 CAGAGGGGTGGGACAGGGGCGGGGGTTCCCGCTGTATACCATCCCTG-CCATACCAAC 660  
|||||  
QY 1987 CCAGTATTAATCTCGCTGGTTTCTGTTTATTTAAT 2025  
|||||  
Db 661 CCAGTATTAATCTCGCATGTTTCTGCTGTTTCAAT 699  
|||||

RESULT 15  
BM790595  
LOCUS  
DEFINITION K-EST0070459 S21SN520 Homo sapiens cDNA clone S21SN520-11-G11 5',  
mRNA sequence.  
ACCESSION BM790595  
VERSION BM790595.1 GI:19138827  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 646)  
AUTHORS Oh,K.J., Cheong,J.E., Sohn,H.Y., Kim,J.M., Park,H.S., Kim,S. and  
Kim,Y.S.  
TITLE 21C Frontier Korean EST Project 2001  
JOURNAL Unpublished (2002)  
COMMENT Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470

Fax: +82-42-860-4409  
Email: Yongsung@mail.kribb.re.kr  
Plate: 11 row: G column: 11  
High quality sequence stop: 646.  
Location/Qualifiers  
1. 646  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="S21SN520-11-G11"  
/clone\_lib="S21SN520"  
/sex="F"  
/tissue\_type="Stomach"  
/cell\_type="Floating aggregates"  
/cell\_line="SNU-520"  
/lab\_host="Top10F"  
/note="Organ: Stomach; Vector: pTZ18RP1; Site\_1: EcoRI;  
Site\_2: NotI; The poly (A)+ RNA was dephosphorylated with  
bacterial alkaline phosphatase (BAP) and then decapped  
with tobacco acid pyrophosphatase (TAP). The decapped  
intact mRNA was ligated with DNA-RNA linker including EcoR  
I site by treatment of T4 RNA ligase and the first strand  
cDNA was synthesized from oligo dt-selected mRNA by  
priming with dt-tailed vector. The dt-tailed vector was  
adjusted to have about 60nt. The cDNA vector was  
circularized with E. coli DNA ligase after digestion of  
EcoRI which site is also included in vector. An RNA strand  
converted to a DNA strand by Okayama-Berg method. The  
obtained cDNA vectors were used for transformation of  
competent cells E. coli Top10F' by electroporation method.  
The cDNA libraries constructed by this method are  
full-length enriched cDNA library."  
BASE COUNT 140 a 213 c 184 g 109 t  
ORIGIN  
Query Match 30.4%; Score 635; DB 14; Length 646;  
Best Local Similarity 99.8%; Pred. No. 1.2e-80;  
Matches 646; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
QY 654 GCGGGAGTGGGGAGCTCATTTGAGAAGTGGCGAAAGCGCACGAAACCTTCCCTGC 713  
|||||  
Db 1 GCGGGAGTGGGGAGCTCATTTGAGAAGTGGCGAAAGCGCACGAAACCTTCCCTGC 60  
|||||  
QY 714 CTTCTGCAGCTGGGAAATACACTACGAAACAGCTCAGAAACAGTGTCTCTGGA 773  
|||||  
Db 61 CTTCTG-CAGCTGGGAAATACACTACGAAACAGCTCAGAAACAGTGTCTCTGGA 119  
|||||  
QY 774 CATTGACCTCTGGGACAAAGTTTCACTGAACTCTCCACCAAGTGCATCTTAAGACTGTGGA 833  
|||||  
Db 120 CATTGACCTCTGGGACAAAGTTTCACTGAACTCTCCACCAAGTGCATCTTAAGACTGTGGA 179  
|||||  
QY 834 GTTGGCAAGCAGCTGCCGGCTTACCCACCTCACCATCGCCGACGAGTACCCCTCT 893  
|||||  
Db 180 GTTGGCAAGCAGCTGCCGGCTTACCCACCTCACCATCGCCGACGAGTACCCCTCT 239  
|||||  
QY 894 CRAAGGTGCTGCTGGACATCTCTGATCTCTGGGATGTGACGGGTACACGCCCGAGCA 953  
|||||  
Db 240 CRAAGGTGCTGCTGGACATCTCTGATCTCTGGGATGTGACGGGTACACGCCCGAGCA 299  
|||||  
QY 954 GGACACCATGACCTTCTCGGACGGGTGACCTTGAACGCCGACGAGTGCACACGCTGG 1013  
|||||  
Db 300 GGACACCATGACCTTCTCGGACGGGTGACCTTGAACGCCGACGAGTGCACACGCTGG 359  
|||||  
QY 1014 CTTGGCCCCCTCACCAGCTTGTGCTTTTGGCTTTCGCCAACACAGCTGTGCCCTTGGAGAT 1073  
|||||  
Db 360 CTTGGCCCCCTCACCAGCTTGTGCTTTTGGCTTTCGCCAACACAGCTGTGCCCTTGGAGAT 419  
|||||  
QY 1074 GGATGATGGGAGACGGGGTGTCTAGGCCCATCTGCTCTATCTGCGGAGACGCCAGGA 1133  
|||||  
Db 420 GGATGATGGGAGACGGGGTGTCTAGGCCCATCTGCTCTATCTGCGGAGACGCCAGGA 479  
|||||  
QY 1134 CTTGGAGACGGGACCGGCTGGGATGATCTGTGAGGAGCCGCTGCTGGAGGCGCTAAAGGT 1193  
|||||  
Db 480 CTTGGAGACGGGACCGGCTGGGATGATCTGTGAGGAGCCGCTGCTGGAGGCGCTAAAGGT 539  
|||||



Result No.	Query			DB	ID	Description
	Score	Match	Length			
1	1782.2	85.4	3036	1	US-08-306-691B-52	Sequence 52, Appl
2	1782.2	85.4	3036	2	US-08-095-728B-1	Sequence 1, Appl
3	1782.2	85.4	3036	5	PCF-US92-02320A-1	Sequence 1, Appl
4	1782.2	85.3	3511	3	US-08-892-747-13	Sequence 13, Appl
5	1778.8	85.3	2928	2	US-08-095-728B-3	Sequence 3, Appl
6	1778.8	85.3	2928	5	PCF-US92-02320A-3	Sequence 3, Appl
7	1777.2	85.2	2940	2	US-08-592-383-1	Sequence 1, Appl
8	1777.2	85.2	2840	6	517671-1	Patent No. 517671
9	1200	57.5	2658	2	US-08-592-383-3	Sequence 3, Appl
10	749.2	35.9	1576	6	5260432-1	Patent No. 5260432
11	702.4	33.7	704	2	US-08-592-383-5	Sequence 5, Appl
12	667.8	32.0	2989	6	5223606-1	Patent No. 5223606
13	293.4	14.1	558	2	US-08-896-365-5	Sequence 5, Appl
14	199.4	9.6	201	2	US-08-592-383-6	Sequence 6, Appl
15	183	8.8	1866	1	US-08-336-408B-1	Sequence 1, Appl
16	183	8.8	1866	5	PCF-US91-00399-1	Sequence 1, Appl
17	180.2	8.6	1893	6	5438136-1	Patent No. 5438136
18	177.6	8.5	2095	1	US-08-336-408B-3	Sequence 3, Appl
19	177.6	8.5	2095	5	PCF-US91-00399-3	Sequence 3, Appl
20	160.8	7.7	2204	1	US-07-952-800-3	Sequence 3, Appl
21	160.8	7.7	2219	1	US-08-336-408B-7	Sequence 7, Appl
22	160.8	7.7	2219	5	PCF-US91-00399-7	Sequence 7, Appl
23	159.6	7.7	816	1	US-08-485-971-22	Sequence 22, Appl
24	159.6	7.7	816	1	US-08-383-754-22	Sequence 22, Appl
25	159.6	7.7	816	1	US-08-485-978-22	Sequence 22, Appl
26	159.6	7.7	816	2	US-08-486-814-22	Sequence 22, Appl
27	159.6	7.7	816	2	US-08-487-472-22	Sequence 22, Appl



FILED DATE: 21-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/673,838  
FILING DATE: 22-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: WHITE, JOHN P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 38694-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 3036 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
CLONE: MYL-RAR  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 67..2457  
OTHER INFORMATION:  
US-08-095-728B-1

Query Match 85.4%; Score 1782.2; DB 2; Length 3036;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1784; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 290 CCATTGAGACCCAGAGCAGCAGTTCCTGAAGAGATAGTGCACAGCCCTCCCTCGCCACCC 349  
DB 1250 CCATTGAGACCCAGAGCAGCAGTTCCTGAAGAGATAGTGCACAGCCCTCCCTCGCCACCC 1309  
QY 350 CTCCTACCCCGCATCTACAGCGCTTCTTGTCTGTCTGAGGACAGTCTCAGGCTACCACT 409  
DB 1310 CTCCTACCCCGCATCTACAGCGCTTCTTGTCTGTCTGAGGACAGTCTCAGGCTACCACT 1369  
QY 410 ATGGGGTCAGCGCCCTGTGAGGGCTCAAGGGCTTCTTCCGCGCAGCAGTCCAGAGAACA 469  
DB 1370 ATGGGGTCAGCGCCCTGTGAGGGCTCAAGGGCTTCTTCCGCGCAGCAGTCCAGAGAACA 1429  
QY 470 TGGGTACAGCTGTACCGGGACAGAGTGCATCATCAACAGGTGACCCCGGAACCCCT 529  
DB 1430 TGGGTACAGCTGTACCGGGACAGAGTGCATCATCAACAGGTGACCCCGGAACCCCT 1489  
QY 530 GCCAGTACTGCCGACTGCAGAGTCTTGAAGTGGGCATGTCCAGGAGTCTGTGAGAA 589  
DB 1490 GCCAGTACTGCCGACTGCAGAGTCTTGAAGTGGGCATGTCCAGGAGTCTGTGAGAA 1549  
QY 590 ACGACCGAAACAGAGAGAGGTGCCCCAAGCCGAGTGTCTGAGAGCTACACGC 649  
DB 1550 ACGACCGAAACAGAGAGAGGTGCCCCAAGCCGAGTGTCTGAGAGCTACACGC 1609  
QY 650 TGACCCCGAGGTGGGGAGTCAATGAGAGGTGGGCAAGCGACCAAGGAAACCTTCC 709  
DB 1610 TGACCCCGAGGTGGGGAGTCAATGAGAGGTGGGCAAGCGACCAAGGAAACCTTCC 1669  
QY 710 CTGCCCTCTGCAGCTGGGCAATACATACCAACAGCTCAGAACAGCTGTCTCTC 769  
DB 1670 CTGCCCTCTGCAGCTGGGCAATACATACCAACAGCTCAGAACAGCTGTCTCTC 1729  
QY 770 TGGACATTGACCTCTGGGACAGTTCAGTGAATCTCCACCAAGTGCATTAAGACTG 829  
DB 1730 TGGACATTGACCTCTGGGACAGTTCAGTGAATCTCCACCAAGTGCATTAAGACTG 1789  
QY 830 TGGAGTTCCCAAGCAGTGGCCGCTTACACACCTTCCATCGCCGACCAAGTACACC 889  
DB 1790 TGGAGTTCCCAAGCAGTGGCCGCTTACACACCTTCCATCGCCGACCAAGTACACC 1849  
QY 890 TCCTCAAGGCTGCCTGCTGACATCTCTGATCTCGGATCTGCAGCGGTACAGCGCG 949  
|||||

DB 1850 TCCTCAAGGCTGCTGCTGGACATCTCTGATCTCGGATCTGCACCGGGTACAGCCGC 1909  
QY 950 AGCAGACACCATGACCTTCTCGAGCGGCTGACCTTAACCGGACCCAGATGACACAACG 1009  
DB 1910 AGCAGACACCATGACCTTCTCGAGCGGCTGACCTTAACCGGACCCAGATGACACAACG 1969  
QY 1010 CTGGCTTGGCCCGCTTACCGACCTGCTTCTGCTTTCGCTTTCGCAACAGCTGCTGCCCTGG 1069  
DB 1970 CTGGCTTGGCCCGCTTACCGACCTGCTTCTGCTTTCGCTTTCGCAACAGCTGCTGCCCTGG 2029  
QY 1070 AGATGATGATGCGGAGACGGGCTGCTCAGCGCCATCTGCCTCATCTGCGGAGACCGCC 1129  
DB 2030 AGATGATGATGCGGAGACGGGCTGCTCAGCGCCATCTGCCTCATCTGCGGAGACCGCC 2089  
QY 1130 AGGACCTGAGCAGCCCGGCTGACATGCTGACGAGCGCTGCTGAGAGCGCTAA 1189  
DB 2090 AGGACCTGAGCAGCCCGGCTGACATGCTGACGAGCGCTGCTGAGAGCGCTAA 2149  
QY 1190 AGGCTTACGTCGGAAGCGGAGCGCCGAGCGCCGAGCGCTGATCAGCTGAAGA 1249  
DB 2150 AGGCTTACGTCGGAAGCGGAGCGCCGAGCGCGCTGATCAGCTGAAGA 2209  
QY 1250 AGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGAGCGGCTGATCAGCTGAAGA 1309  
DB 2210 AGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGAGCGGCTGATCAGCTGAAGA 2269  
QY 1310 TGGAGATCCCGGCTCCTATGCCCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGGCC 1369  
DB 2270 TGGAGATCCCGGCTCCTATGCCCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGGCC 2329  
QY 1370 TGGAGACTCTGAGCGGACAGCCGGGGGTGGGGGGGAGGGGGTGGCTGCCCCCCC 1429  
DB 2330 TGGAGACTCTGAGCGGACAGCCGGGGGTGGGGGGGAGGGGGTGGCTGCCCCCCC 2389  
QY 1430 CGCCAGCGAGCTGACGCCAGCTCAGCGCCAGCTCCAAAGAGAGCAGCGCCGCGCACCC 1489  
DB 2390 CGCCAGCGAGCTGACGCCAGCTCAGCGCCAGCTCCAAAGAGAGCAGCGCCGCGCACCC 2449  
QY 1490 ACTCCCGCTGACCGCCAGCCACATGGACACAGCCCTCGCCCTCCCGCCCGCTTCT 1549  
DB 2450 ACTCCCGCTGACCGCCAGCCACATGGACACAGCCCTCGCCCTCCCGCCCGCTTCT 2509  
QY 1550 CTGGCTTTTACCGACCATGTGACCCCGCAGCGCTGCCCGACCTGCCCCCTCCCGGGC 1609  
DB 2510 CTGGCTTTTACCGACCATGTGACCCCGCAGCGCTGCCCGACCTGCCCCCTCCCGGGC 2569  
QY 1610 AGTACTGGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGCGGCTCTTGACAGAG 1669  
DB 2570 AGTACTGGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGCGGCTCTTGACAGAG 2629  
QY 1670 GCTGGGGCTCAGTGGACTGCTGCTCCACAGCTGGGCTGACGTCAGAGGCGGAGGC 1729  
DB 2630 GCTGGGGCTCAGTGGACTGCTGCTCCACAGCTGGGCTGACGTCAGAGGCGGAGGC 2689  
QY 1730 CAGGAACTGAGTGAGCGCCCTGCTGCTGGGTCTCAGGATGGGTCTGGGGGCTCTGCTT 1789  
DB 2690 CAGGAACTGAGTGAGCGCCCTGCTGCTGGGTCTCAGGATGGGTCTGGGGGCTCTGCTT 2749  
QY 1790 CATCAGACACCCCTTGCCTCCAGCTCACCACATCTTCATCACCAGCAAGCGGAGCTT 1849  
DB 2750 CATCAGACACCCCTTGCCTCCAGCTCACCACATCTTCATCACCAGCAAGCGGAGCTT 2809  
QY 1850 GGTCTCCCTCCTCAGAACTCACAAGCATTTGCTCCCAAGCTGGGAACTCAACCTCC 1909  
DB 2810 GGTCTCCCTCCTCAGAACTCACAAGCATTTGCTCCCAAGCTGGGAACTCAACCTCC 2869  
QY 1910 CCCCTGCTCGGTGGTGACAGAGGGGTGGGACAGGGCGGGGGTTCCTCCCTGTATAT 1969  
DB 2870 CCCCTGCTCGGTGGTGACAGAGGGGTGGGACAGGGCGGGGGTTCCTCCCTGTATAT 2929  
QY 1970 ACCCTGCCATACCAACCCAGGTATTAATTCGCTGGTGTCTTCTTTATTTATTTT 2029  
DB 2930 ACCCTGCCATACCAACCCAGGTATTAATTCGCTGGTGTCTTCTTTATTTATTTT 2989

QY 2030 TTGTTTGGATTTTAAATAAGAAATTTTCATTTTAAAGCACAATAA 2076  
|||||  
Db 2990 TTGTTTGGATTTTAAATAAGAAATTTTCATTTTAAAGCACAATAA 3036

## RESULT 3

PCT-US92-02320A-1  
; Sequence 1, Application PC/TUS9202320A  
; GENERAL INFORMATION:  
; APPLICANT: Sloan-Kettering Institute, For Cancer Research  
; TITLE OF INVENTION: METHODS FOR DETECTION AND TREATMENT OF CANCER  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOPER & DUNHAM  
; STREET: 30 ROCKEFELLER PLAZA  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/02320A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 675,084  
; FILING DATE: 22-MAR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 673,838  
; FILING DATE: 22-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WHITE, JOHN P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 38694-PCT  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 977-9550  
; TELEFAX: (212) 644-0525  
; TELEX: (212) 422523 COOP UI  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3036 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; IMMEDIATE SOURCE:  
; CLONE: MYL-RAR  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 67..2457  
; OTHER INFORMATION:  
PCT-US92-02320A-1

Query Match 85.4%; Score 1782.2; DB 5; Length 3036;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1784; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 290 CCATTGAGACCCAGACAGAGTCTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACCCC 349  
|||||  
Db 1250 CCATTGAGACCCAGACAGAGTCTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACCCC 1309  
|||||  
QY 350 CTCTACCCCGCATCTACAAGCCTTGCTTCTCTCAGGACAAGTCTCTCAGGCTACCACT 409  
|||||  
Db 1310 CTCTACCCCGCATCTACAAGCCTTGCTTCTCTCAGGACAAGTCTCTCAGGCTACCACT 1369  
|||||  
QY 410 ATGGGGTACGCGCTGTGAGGGGTGCAAGGGCTTCTTCCGCGCAGCATCCAGAAGAACA 469  
|||||  
Db 1370 ATGGGGTACGCGCTGTGAGGGGTGCAAGGGCTTCTTCCGCGCAGCATCCAGAAGAACA 1429  
|||||

QY 470 TGTGTACACGTGTACCCGGGCAAGAACTGCATCATCAACAAAGGTGACCCGGAACCCCT 529  
|||||  
Db 1430 TGTGTACACGTGTACCCGGGCAAGAACTGCATCATCAACAAAGGTGACCCGGAACCCCT 1489  
|||||  
QY 530 GCCAGTACTGCCGACTGCAGAAAGTGTCTTGAAGTGGGCATGTCTCAAGAGTCTGTGAGAA 589  
|||||  
Db 1490 GCCAGTACTGCCGACTGCAGAAAGTGTCTTGAAGTGGGCATGTCTCAAGAGTCTGTGAGAA 1549  
|||||  
QY 590 ACAGCCGAACAAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTCTGAGAGTCTACAGCG 649  
|||||  
Db 1550 ACAGCCGAACAAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTCTGAGAGTCTACAGCG 1609  
|||||  
QY 650 TGACGCCGAGGTGGGGAGTCTATTGAGAAAGTGGCAAGCGCACAGAGAACTTCC 709  
|||||  
Db 1610 TGACGCCGAGGTGGGGAGTCTATTGAGAAAGTGGCAAGCGCACAGAGAACTTCC 1669  
|||||  
QY 710 CTGCCCTCTGCCAGCTGGGCAATACACTACGACACACAGCTCAGAACACAGTGTCTCTC 769  
|||||  
Db 1670 CTGCCCTCTGCCAGCTGGGCAATACACTACGACACACAGCTCAGAACACAGTGTCTCTC 1729  
|||||  
QY 770 TGGACATTTGACCTCTCGGACAACTTCAGTGAACCTCTCCACCAAGTGCATCATTAAGACTG 829  
|||||  
Db 1730 TGGACATTTGACCTCTCGGACAACTTCAGTGAACCTCTCCACCAAGTGCATCATTAAGACTG 1789  
|||||  
QY 830 TGGAGTTGCCCAAGCAGCTGCCCGGCTTCACCACTCCACCTCAGCCGACAGATCACCC 889  
|||||  
Db 1790 TGGAGTTGCCCAAGCAGCTGCCCGGCTTCACCACTCCACCTCAGCCGACAGATCACCC 1849  
|||||  
QY 890 TCCTCAAGCTGCTGCCCTGGACATCTCTGCGGATCTGCACGCGGTACACGCGCG 949  
|||||  
Db 1850 TCCTCAAGCTGCTGCCCTGGACATCTCTGCGGATCTGCACGCGGTACACGCGCG 1909  
|||||  
QY 950 AGCAGGACACCATGACCTTCTCGGACGGCTGACCTCTGAACCGGACCCAGATGCACAAACG 1009  
|||||  
Db 1910 AGCAGGACACCATGACCTTCTCGGACGGCTGACCTCTGAACCGGACCCAGATGCACAAACG 1969  
|||||  
QY 1010 CTGGCTTGGCGCCCTCAGCAGCTGGTCTTTGGCTTTCGCAACCAAGTGTCTGCCCTGG 1069  
|||||  
Db 1970 CTGGCTTGGCGCCCTCAGCAGCTGGTCTTTGGCTTTCGCAACCAAGTGTCTGCCCTGG 2029  
|||||  
QY 1070 AGATGGATGATCGGAGACGGGCTGCTCAGCGCCATCTGCCTCATCTCGGGAGACGGCC 1129  
|||||  
Db 2030 AGATGGATGATCGGAGACGGGCTGCTCAGCGCCATCTGCCTCATCTCGGGAGACGGCC 2089  
|||||  
QY 1130 AGGACCTGGAGCAGCGGACCGGGTGGACATGCTGCAGGAGCGGCTGCTGGAGCGCGTAA 1189  
|||||  
Db 2090 AGGACCTGGAGCAGCGGACCGGGTGGACATGCTGCAGGAGCGGCTGCTGGAGCGCGTAA 2149  
|||||  
QY 1190 AGGTCTACGTGGGAAGCGGAGGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGA 1249  
|||||  
Db 2150 AGGTCTACGTGGGAAGCGGAGGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGA 2209  
|||||  
QY 1250 AGATTACTGACCTGCGAAGCATCAGCGCAAGGGGCTGAGGGGTGATCAGCTGAAGA 1309  
|||||  
Db 2210 AGATTACTGACCTGCGAAGCATCAGCGCAAGGGGCTGAGGGGTGATCAGCTGAAGA 2269  
|||||  
QY 1310 TGGAGATCCCGGGCTCCATCGCGCTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCC 1369  
|||||  
Db 2270 TGGAGATCCCGGGCTCCATCGCGCTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCC 2329  
|||||  
QY 1370 TGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGAGCGGGGTGCGCTGCCCGCCC 1429  
|||||  
Db 2330 TGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGAGCGGGGTGCGCTGCCCGCCC 2389  
|||||  
QY 1430 CGCCAGGAGCTGTAGCCCCAGCTCAGCCCCAGCTCCAAACAGAGACGCCCGGCCACCC 1489  
|||||  
Db 2390 CGCCAGGAGCTGTAGCCCCAGCTCAGCCCCAGCTCCAAACAGAGACGCCCGGCCACCC 2449  
|||||  
QY 1490 ACTCCCGTGACCGCCACCGCCACATGACACAGCCCTCGCCCTCGCCCGCGCTTTTCT 1549  
|||||  
Db 2450 ACTCCCGTGACCGCCACCGCCACATGACACAGCCCTCGCCCTCGCCCGCGCTTTTCT 2509  
|||||



Qy	1550	CTGCGTTTCTTACGACATGTGACCCCGACAGCCCTGCGCCCGACACTGCGCCTCCCGGCG	1609
Db	2510	CTGCGTTTCTTACGACCATGTGACCCCGCACAGCCCTGCGCCACCTGCGCCTCCCGGCG	2569
Qy	1610	AGTACTGGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGCAGCGCACTCCTTGGACAGAG	1669
Db	2570	AGTACTGGGGACCTTCCCTGGGGAGCGGGAGGAGGAGCGGACACTCTTGGACAGAG	2629
Qy	1670	GCCTGGGCGCCTCAGTGGACTGCTGCTCCACAGAGCCTGGGCTGACGTGAGAGGCCGAGGC	1729
Db	2630	GCCTGGGCGCCTCAGTGGACTGCTGCTCCACAGAGCCTGGGCTGACGTGAGAGGCCGAGGC	2689
Qy	1730	CAGGAATGAGTGAGGCGCCCTGGTCTCTGGGTCTCAGGATGGTCTCGGGGCGCTCGTGT	1789
Db	2690	CAGGAATGAGTGAGGCGCCCTGGTCTCTGGGTCTCAGGATGGTCTCGGGGCGCTCGTGT	2749
Qy	1790	CATCAGACACCCCTCTGCCGAGCTCACCACATCTTCATCACCAGCAAAAGCCAGGACTT	1849
Db	2750	CATCAGACACCCCTCTGCCGAGCTCACCACATCTTCATCACCAGCAAAAGCCAGGACTT	2809
Qy	1850	GGCTCCCGCATCTCAGAACTCACAGGCCATTGCTCCCGACCTGGGGAACCTCAACCTCC	1909
Db	2810	GGCTCCCGCATCTCAGAACTCACAGGCCATTGCTCCCGACCTGGGGAACCTCAACCTCC	2869
Qy	1910	CCCTGCGCTCGTGTGTGACAGAGGGGTGGGACAGGGCGGGGGGTTCGCCCTGTACAT	1969
Db	2870	CCCTGCGCTCGTGTGTGACAGAGGGGTGGGACAGGGCGGGGGGTTCGCCCTGTACAT	2929
Qy	1970	ACCTGCCATACAAACCCAGGTATTAATTCCTCGCTGCTTTGTTTTTATTTTAAATTTT	2029
Db	2930	ACCTGCCATACAAACCCAGGTATTAATTCCTCGCTGCTTTGTTTTTATTTTAAATTTT	2989
Qy	2030	TTGTTTTGATTTTTTAAATGAATTTTCATTTTAAGCACAATAAAA	2076
Db	2990	TTGTTTTGATTTTTTAAATGAATTTTCATTTTAAGCACAATAAAA	3036

## RESULT 4

RESOL 4  
US-08-892-747-13  
; Sequence 13, Application US/08892747  
; Patent No. 6057153  
; GENERAL INFORMATION:  
; APPLICANT: Shaji T. George, Michael Ma, Martina Werner,  
; APPLICANT: Umberto Pace and Allan R. Goldberg  
; TITLE OF INVENTION: Stabilized External Guide Sequences  
; NUMBER OF SEQUENCES: 50  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Patrea L. Pabst  
; STREET: 2800 One Atlantic Center  
; STREET: 1201 West Peachtree Street  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30309-3450  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA: #1.0, Version #1.25  
; APPLICATION NUMBER: US/08/892,747  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/372,556  
; FILING DATE: January 13, 1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US96/00513  
; FILING DATE: January 19, 1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Pabst, Patrea L.  
;



Db 2445 CTGGCTTGGGCGCCCTACCGACCTGGTCTTGGCTTGGCAACACAGCTGCTGCCCCG 2504  
QY 1070 AGATGGATGATCGGAGACGGGCTGCTCAGCGCATCTGCTCATCTCGGAGACGCC 1129  
Db 2505 AGATGGATGATCGGAGACGGGCTGCTCAGCGCATCTGCTCATCTCGGAGACGCC 2564  
QY 1130 AGGACCTGGAGACCGGACCGGGTGGACATGCTCGAGGACCGCTGCTGGAGGCGTAA 1189  
Db 2565 AGGACCTGGAGACCGGACCGGGTGGACATGCTCGAGGACCGCTGCTGGAGGCGTAA 2624  
QY 1190 AGGCTACGTCGGAAGGAGGCGGACCGGCGCCCGCCACATGTTCCCAAGATGCTAATGA 1249  
Db 2625 AGGCTACGTCGGAAGGAGGCGGACCGGCGCCCGCCACATGTTCCCAAGATGCTAATGA 2684  
QY 1250 AGATTACTGACCTCGAAGCATCAGCGCCCAAGGGGCTGAGCGGTGATCAGCTGAAGA 1309  
Db 2685 AGATTACTGACCTCGAAGCATCAGCGCCCAAGGGGCTGAGCGGTGATCAGCTGAAGA 2744  
QY 1310 TGGAGATCCCGGGTCCATGCCGCTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCC 1369  
Db 2745 TGGAGATCCCGGGTCCATGCCGCTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCC 2804  
QY 1370 TGGACACTCTGAGGGACACCGGGGGTGGGGGGCGGAGCGGGGTGGCTGSCCCCC 1429  
Db 2805 TGGACACTCTGAGGGACACCGGGGGTGGGGGGCGGAGCGGGGTGGCTGSCCCCC 2864  
QY 1430 CGCAGGACGCTGAGCGCCACGCTCAGCGCCAGCTCAGAGAGAGAGCGCGGCCACCC 1489  
Db 2865 CGCAGGACGCTGAGCGCCACGCTCAGCGCCAGCTCAGAGAGAGAGCGCGGCCACCC 2924  
QY 1490 ACTCCCGTGGACCGCCACGCTGAGGACACAGCGCTCGCCCTCCGCGCGGCTTTCT 1549  
Db 2925 ACTCCCGTGGACCGCCACGCTGAGGACACAGCGCTCGCCCTCCGCGCGGCTTTCT 2984  
QY 1550 CTGCTTTTACCGACATGTGACCCCGCACAGCGCTGCCCCACCTGCGCTCCCGGGC 1609  
Db 2985 CTGCTTTTACCGACATGTGACCCCGCACAGCGCTGCCCCACCTGCGCTCCCGGGC 3044  
QY 1610 AGTACTGGGACCTTCCCTGGGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1669  
Db 3045 AGTACTGGGACCTTCCCTGGGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3104  
QY 1670 GCTGGGCGCTCAGTGGACTGCTGCTGCTCCACAGCTGGGCTGAGGAGGAGGAGGAGG 1729  
Db 3105 GCTGGGCGCTCAGTGGACTGCTGCTGCTCCACAGCTGGGCTGAGGAGGAGGAGGAGG 3164  
QY 1730 CAGGAACCTGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1789  
Db 3165 CAGGAACCTGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3224  
QY 1790 CATCAGACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1849  
Db 3225 CATCAGACACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3284  
QY 1850 GCTCCCGCATCTCAGACTCAGAGCCATGCTCCCGAGCTGGGAGGAGGAGGAGGAGG 1909  
Db 3285 GCTCCCGCATCTCAGACTCAGAGCCATGCTCCCGAGCTGGGAGGAGGAGGAGGAGG 3344  
QY 1910 CCCCTGCTCGTGGTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1969  
Db 3345 CCCCTGCTCGTGGTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3404  
QY 1970 ACCCTGCCATACCAACCCAGTATTAATCTCGTGGTGGTGGTGGTGGTGGTGGTGG 2029  
Db 3405 ACCCTGCCATACCAACCCAGTATTAATCTCGTGGTGGTGGTGGTGGTGGTGGTGG 3464  
QY 2030 TTGTTTTGATTTTTTAAGAAATTTTCAATTTTAAGCAAAAAA 2076  
Db 3465 TTGTTTTGATTTTTTAAGAAATTTTCAATTTTAAGCAAAAAA 3511

RESULT 5

US-08-095-728B-3  
; Sequence 3, Application US/08095728B  
; Patent No. 5843642  
; GENERAL INFORMATION:  
; APPLICANT: DMITROVSKY, ETHAN  
; APPLICANT: WARRELL JR, RAYMOND P  
; APPLICANT: MILLER JR, WILSON H  
; APPLICANT: FRANKEL, STANLEY  
; TITLE OF INVENTION: METHODS FOR THE DETECTION AND  
; TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA (APL)  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOPER & DUNHAM LLP  
; STREET: 1185 AVENUE OF THE AMERICAS  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/095,728B  
; FILING DATE: 21-JUL-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/673,838  
; FILING DATE: 22-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WHITE, JOHN P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 38694-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2928 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; IMMEDIATE SOURCE:  
; CLONE: hRAR ALPHA  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 103..1488  
; OTHER INFORMATION:  
US-08-095-728B-3

Query Match 85.3%; Score 1778.8; DB 2; Length 2928;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 1786; Conservative 0; Mismatches 12; Indels 0; Gaps 0;

QY 288 CTCATTGAGACCCAGAGCAGCAGCTTCTGAAGAGATAGTCCAGAGCCTCCCTCGCCACC 347  
Db 279 CACCATGAGACCCAGAGCAGCAGCTTCTGAAGAGATAGTCCAGAGCCTCCCTCGCCACC 338  
QY 348 CCCTCTACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAAGTCTCTCAGGCTACCA 407  
Db 339 CCCTCTACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAAGTCTCTCAGGCTACCA 398  
QY 408 CTATGGGGTCAGGCGCTGTGAGGGCTGCAAGGCTTCTTCCCGCGCAGCATCCAGAGAA 467  
Db 399 CTATGGGGTCAGGCGCTGTGAGGGCTGCAAGGCTTCTTCCCGCGCAGCATCCAGAGAA 458  
QY 468 CATGGTGTACAGTGTTCACCGGGACAGAACTGCATCATCAACAAGGTGACCGGAACCC 527  
Db 459 CATGGTGTACAGTGTTCACCGGGACAGAACTGCATCATCAACAAGGTGACCGGAACCC 518  
QY 528 CTGCCAGTACTCCGACTGCAGAAAGTCTTTGAAGTGGGCTATGTCACCAAGGAGTCTGTGAG 587



TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
IMMEDIATE SOURCE:  
CLONE: hRAR ALPHA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 103..1488  
OTHER INFORMATION:  
PCT-US92-02320A-3

Query Match		85.3%	Score 1778.8;	DB 5;	Length 2928;
Best Local Similarity		99.3%	Pred. No. 0;		
Matches 1786;		Conservative 0;	Mismatches 12;	Indels 0;	Gaps 0;
QY	288	CTCATTTGAGACCCAGACAGCAGCTTCTGAAGAGATAGTGCACAGCCTCCCTCGCCACC	347		
DB	279	CACCATTTGACCCAGACAGCAGCTTCTGAAGAGATAGTGCACAGCCTCCCTCGCCACC	338		
QY	348	CCCTCTACCCCGCATCTACAAAGCCTTGTCTGTCTGTCAGGACAAGTCTCAGGCTACCA	407		
DB	339	CCCTCTACCCCGCATCTACAAAGCCTTGTCTGTCTGTCAGGACAAGTCTCAGGCTACCA	398		
QY	408	CTATGGGGTCAGCGCCTGTGAGGGCTGCAAGGGCTTCTCCGCGCAGCATCCAGAGAA	467		
DB	399	CTATGGGGTCAGCGCCTGTGAGGGCTGCAAGGGCTTCTCCGCGCAGCATCCAGAGAA	458		
QY	468	CATGGTGTACACGTGTACCCGGGACAGAACTGCATCATCAACAAAGTGACCCGGAAACC	527		
DB	459	CATGGTGTACACGTGTACCCGGGACAGAACTGCATCATCAACAAAGTGACCCGGAAACC	518		
QY	528	CTGCCAGTACTGCGAGTGTGAGAGTGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAG	587		
DB	519	CTGCCAGTACTGCGAGTGTGAGAGTGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAG	578		
QY	588	AAACGACCGAAACAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTGAGAGCTACAC	647		
DB	579	AAACGACCGAAACAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTGAGAGCTACAC	638		
QY	648	GCTGACCGCGAGGTGGGGAGCTCATTTGAGAGAGTGCAGAAAGCGCACCAGAAACCTT	707		
DB	639	GCTGACCGCGAGGTGGGGAGCTCATTTGAGAGAGTGCAGAAAGCGCACCAGAAACCTT	698		
QY	708	CCCTGCCCTCTGCGAGTGGGCAAAATACACTAGGAAACACAGCTCAGAAACAGTGTCTC	767		
DB	699	CCCTGCCCTCTGCGAGTGGGCAAAATACACTAGGAAACACAGCTCAGAAACAGTGTCTC	758		
QY	768	TCTGGACATGACCTCTGGGACAAAGTTCAGTGAATCTCCACCAAGTGCATATTAGAC	827		
DB	759	TCTGGACATGACCTCTGGGACAAAGTTCAGTGAATCTCCACCAAGTGCATATTAGAC	818		
QY	828	TGTGAGTTTCGCGAAGCAGCTGCCGGCTTCACCAACCTCACCATGCCGACAGATCAC	887		
DB	819	TGTGAGTTTCGCGAAGCAGCTGCCGGCTTCACCAACCTCACCATGCCGACAGATCAC	878		
QY	888	CTCTCTCAAGGCTGCTGCTGGACATCTCTGATCTCTGCGGATCTGACGCGGTACACGCC	947		
DB	879	CTCTCTCAAGGCTGCTGCTGGACATCTCTGATCTCTGCGGATCTGACGCGGTACACGCC	938		
QY	948	CGAGCAGGACACCATGACCTTCTCGACCGGGTGCACCTGAAACCGGACCCAGATGCACAA	1007		
DB	939	CGAGCAGGACACCATGACCTTCTCGACCGGGTGCACCTGAAACCGGACCCAGATGCACAA	998		
QY	1008	CGCTGGCTTTCGCGCCCTCACCAGACCTGTCTTTGCTTTCGCAACACAGCTGCGCCCT	1067		
DB	999	CGCTGGCTTTCGCGCCCTCACCAGACCTGTCTTTGCTTTCGCAACACAGCTGCGCCCT	1058		
QY	1068	GGAGATGGATGATGCGGAGACGGGGTGTCTGACGCGCATCTGCCTCATCTGCGGAGACCG	1127		
DB	1059	GGAGATGGATGATGCGGAGACGGGGTGTCTGACGCGCATCTGCCTCATCTGCGGAGACCG	1118		
QY	1128	CCAGGACCTGGAGACCGGACCGGGTGGACATGTCTGACGAGCCGCTGCTGAGGCGCT	1187		
DB	1119	CCAGGACCTGGAGACCGGACCGGGTGGACATGTCTGACGAGCCGCTGCTGAGGCGCT	1178		

QY	1188	AAAGGTCTACGTGCGGAAGCGGAGGCCAGCGCCGCCACATGTTCCCCCAAGATGCTAAT	1247		
DB	1179	AAAGGTCTACGTGCGGAAGCGGAGGCCAGCGCCGCCACATGTTCCCCCAAGATGCTAAT	1238		
QY	1248	GAAGATTACTGACCTGCGAAGCATCAGCGCCAAAGGGCTGAGCGGGTGCATCACGCTGAA	1307		
DB	1239	GAAGATTACTGACCTGCGAAGCATCAGCGCCAAAGGGCTGAGCGGGTGCATCACGCTGAA	1298		
QY	1308	GATGGAGATCCCGGCTCCATGCCGCTTCTCATCCAGAAATGTTGGAGAACTCAGAGGG	1367		
DB	1299	GATGGAGATCCCGGCTCCATGCCGCTTCTCATCCAGAAATGTTGGAGAACTCAGAGGG	1358		
QY	1368	CCTGGACACTCTGAGCGACAGCGGGGGGTGGGGGGGCGGACGCGGGTGGCTGCCGCC	1427		
DB	1359	CCTGGACACTCTGAGCGACAGCGGGGGGTGGGGGGGCGGACGCGGGTGGCTGCCGCC	1418		
QY	1428	CCCCCAGGACGCTGTAGCCCCAGCCTCAGCGCCCAAGCTCCCAACAGAGCAGCCCGGCCAC	1487		
DB	1419	CCCCCAGGACGCTGTAGCCCCAGCCTCAGCGCCCAAGCTCCCAACAGAGCAGCCCGGCCAC	1478		
QY	1488	CCACTCCCCGTGACCGCCACCGCCACATGGACACAGCCCTCGCCCTCGCCCGGCTTTT	1547		
DB	1479	CCACTCCCCGTGACCGCCACCGCCACATGGACACAGCCCTCGCCCTCGCCCGGCTTTT	1538		
QY	1548	CTCTGCCCTTTTACCGACCATGTGACCCCGCACAGCCCTGCCCCACCTGCGCTCCCGG	1607		
DB	1539	CTCTGCCCTTTTACCGACCATGTGACCCCGCACAGCCCTGCCCCACCTGCGCTCCCGG	1598		
QY	1608	GCAGTACTGGGAGCTTCCCTGGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1667		
DB	1599	GCAGTACTGGGAGCTTCCCTGGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1658		
QY	1668	AGGCTTGGGGCTCAGTGGAGTGTCTTCCACAGCTGGGCTGACCTCAGAGGCCGAG	1727		
DB	1659	AGGCTTGGGGCTCAGTGGAGTGTCTTCCACAGCTGGGCTGACCTCAGAGGCCGAG	1718		
QY	1728	GCACGAACTGAGTGGAGGCGCTTGTCTTGGGTCTCAGGATGGGTCTTGGGGGCTCTGTG	1787		
DB	1719	GCACGAACTGAGTGGAGGCGCTTGTCTTGGGTCTCAGGATGGGTCTTGGGGGCTCTGTG	1778		
QY	1788	TTTATCAAGACACCCCTCTGCCAGCTCACCACATCTTATCACCAGCAACGCGCAGGAC	1847		
DB	1779	TTTATCAAGACACCCCTCTGCCAGCTCACCACATCTTATCACCAGCAACGCGCAGGAC	1838		
QY	1848	TTGGCTCCCCCATCTCTCAGAACTCACAAGCCATTGCTCCCCAGCTGGGGAACCTCAACCT	1907		
DB	1839	TTGGCTCCCCCATCTCTCAGAACTCACAAGCCATTGCTCCCCAGCTGGGGAACCTCAACCT	1898		
QY	1908	CCCCCTGCCCTCGGTTGGTGCAGAGGGGGTGGGACAGGGGGGGGGTTCCTCCCTGTAC	1967		
DB	1899	CCCCCTGCCCTCGGTTGGTGCAGAGGGGGTGGGACAGGGGGGGGGTTCCTCCCTGTAC	1958		
QY	1968	ATACCTTGCCTACCAACCCAGGTATTATCTCGCTGTTTGTGTTTATTTTAAATTT	2027		
DB	1959	ATACCTTGCCTACCAACCCAGGTATTATCTCGCTGTTTGTGTTTATTTTAAATTT	2018		
QY	2028	TTTTTGTGTTGATTTTTTAAAGAAATTTTCAATTTAAGCACAAAAAATAAAAAA	2085		
DB	2019	TTTTTGTGTTGATTTTTTAAAGAAATTTTCAATTTAAGCACAAAAAATAAAAAA	2076		

RESULT 7

US-08-592-383-1  
; Sequence 1, Application US/08592383  
; Patent No. 5830760  
; GENERAL INFORMATION:  
; APPLICANT: Tsai, S. and S.J. Collins  
; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acid Res  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue



Qy 1908 CCCCCCTCCCTGGTGTGACAGAGGGGTGGACAGGGGGGGTTCCTCCCTGTAC 1967  
Db 1899 CCCCCCTCCCTGGTGTGACAGAGGGGTGGACAGGGGGGGTTCCTCCCTGTAC 1958  
Qy 1968 ATACCTGCCATACCAACCCAGGATTAATTCCTCGTGGTTCCTTTATTTAAATTT 2027  
Db 1959 ATACCTGCCATACCAACCCAGGATTAATTCCTCGTGGTTCCTTTATTTAAATTT 2018  
Qy 2028 TTTTGGTTTGTATTTTAAATAAGAAATTTTCATTTTAAGCACATTTATCTGAAGGAA 2095  
Db 2019 TTTTGGTTTGTATTTTAAATAAGAAATTTTCATTTTAAGCACATTTATCTGAAGGAA 2076

RESULT 8  
5171671-1  
; Patent No. 5171671  
; APPLICANT: EVANS, RONALD M.; ONG, ESTELITA S.; SEGUI,  
; PRUDIMAR S.; THOMPSON, CATHERINE C.; UEMSONO, KAZUHIKO  
; GUGIERE, VINCENT  
; TITLE OF INVENTION: RETINOIC ACID RECEPTOR COMPOSITION  
; NUMBER OF SEQUENCES: 2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/546,256  
; FILING DATE: 06-AUG-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 276,536  
; FILING DATE: 30-NOV-1988  
; APPLICATION NUMBER: 128,331  
; FILING DATE: 02-DEC-1987  
; SEQ ID NO: 1:  
; LENGTH: 2940  
5171671-1

Query Match 85.2%; Score 1777.2; DB 6; Length 2940;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 1785; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

Qy 288 CTCATTGAGACCCAGACAGCAGTCTCTGAAGAGATAGTCCAGAGCCTCCCTCGCCACC 347  
Db 279 CACCATTTGAGACCCAGACAGCAGTCTCTGAAGAGATAGTCCAGAGCCTCCCTCGCCACC 338  
Qy 348 CCCTCTACCCCGCATCTACAGCCTTCTGTCTGTCAGGACAAAGTCTCAGGCTACCA 407  
Db 339 CCCTCTACCCCGCATCTACAGCCTTCTGTCTGTCAGGACAAAGTCTCAGGCTACCA 398  
Qy 408 CTATGGGTGACGCGCTGTGAGGCTGCAAGGCTTCTTCCGCGGAGCATCCAGAGAA 467  
Db 399 CTATGGGTGACGCGCTGTGAGGCTGCAAGGCTTCTTCCGCGGAGCATCCAGAGAA 458  
Qy 468 CATGGGTGACAGCTGTACCCGGGACAGAACTGCATCATCAACAAGGTGACCCGGAACCC 527  
Db 459 CATGGGTGACAGCTGTACCCGGGACAGAACTGCATCATCAACAAGGTGACCCGGAACCC 518  
Qy 528 CTCCTAGTACTGCCAGTGCAGAGTGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAG 587  
Db 519 CTCCTAGTACTGCCAGTGCAGAGTGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAG 578  
Qy 588 AAACGACCCGAAACAAAGAAAGAGGTGCCCCAGCCGAGTGTCTGAGAGCTACAC 647  
Db 579 AAACGACCCGAAACAAAGAAAGAGGTGCCCCAGCCGAGTGTCTGAGAGCTACAC 638  
Qy 648 GCTGACCGCGGAGGTGGGGAGTCTATTGAGAGGTGCGCAAGCGCACAGGAACCTT 707  
Db 639 GCTGACCGCGGAGGTGGGGAGTCTATTGAGAGGTGCGCAAGCGCACAGGAACCTT 698  
Qy 708 CCTGCCCCCTCTGCCAGCTGGGCAATACACTACGAAACAGCTCAGAAACAGTGTCTC 767  
Db 699 CCTGCCCCCTCTGCCAGCTGGGCAATACACTACGAAACAGCTCAGAAACAGTGTCTC 758  
Qy 768 TCTGGACATTTGACCTCTGGGACAAAGTTTCAAGTGAACCTCTCCACCAAGTGCATTAAGAC 827  
Db 759 TCTGGACATTTGACCTCTGGGACAAAGTTTCAAGTGAACCTCTCCACCAAGTGCATTAAGAC 818

Qy 828 TGTGAGTTTGGCCAAAGCAGCTGCCCGGCTTACCAACCTTACCATCGCCGACAGATCAC 887  
Db 819 TGTGAGTTTGGCCAAAGCAGCTGCCCGGCTTACCAACCTTACCATCGCCGACAGATCAC 878  
Qy 888 CTTCTCAAGGCTGCTGCTGACATCTGATCTGCGGATCTGCACGCGGTACAGCC 947  
Db 879 CTTCTCAAGGCTGCTGCTGACATCTGATCTGCGGATCTGCACGCGGTACAGCC 938  
Qy 948 CGAGCAGGACACCATGACCTTCTCGGACGGGTGACCTTGAACCGGACCCAGATGCACAA 1007  
Db 939 CGAGCAGGACACCATGACCTTCTCGGACGGGTGACCTTGAACCGGACCCAGATGCACAA 998  
Qy 1008 CGCTGGCTTGGCCCCCTTACCGACCTGCTTGGCTTTCGCCAACACAGCTGCTGCCCT 1067  
Db 999 CGCTGGCTTGGCCCCCTTACCGACCTGCTTGGCTTTCGCCAACACAGCTGCTGCCCT 1058  
Qy 1068 GGAGATGATGATCGGAGACGGGCTGCTCAGCGCATCTGCTCAGCGCATCTGCTGCGGAGACCG 1127  
Db 1059 GGAGATGATGATCGGAGACGGGCTGCTCAGCGCATCTGCTCAGCGCATCTGCTGCGGAGACCG 1118  
Qy 1128 CCAGGACCTGGAGCAGCGGCTGACATGCTGAGGAGCGGCTGCTGAGGCGCT 1187  
Db 1119 CCAGGACCTGGAGCAGCGGCTGACATGCTGAGGAGCGGCTGCTGAGGCGCT 1178  
Qy 1188 AAAGTCTACGTGCGGAAGCGGAGCCCGCCAGCCCGCCACATGTTTCCCAAGATGCTAAT 1247  
Db 1179 AAAGTCTACGTGCGGAAGCGGAGCCCGCCAGCCCGCCACATGTTTCCCAAGATGCTAAT 1238  
Qy 1248 GAAGTATTACTGCTGCGAAGCATCAGCGCCAAAGGGGCTGAGCGGGTGTATCACGCTGAA 1307  
Db 1239 GAAGTATTACTGCTGCGAAGCATCAGCGCCAAAGGGGCTGAGCGGGTGTATCACGCTGAA 1298  
Qy 1308 GATGAGATCCCGGCTCCATCGCCCTTCTATCAGAGAAATGTTGGAACTCAGAGGG 1367  
Db 1299 GATGAGATCCCGGCTCCATCGCCCTTCTATCAGAGAAATGTTGGAACTCAGAGGG 1358  
Qy 1368 CTTGGACACTCTGAGCGGACAGCGGGGGGTGGGGGGGAGCGGGGTGCGCTGCCCC 1427  
Db 1359 CTTGGACACTCTGAGCGGACAGCGGGGGGTGGGGGGGAGCGGGGTGCGCTGCCCC 1418  
Qy 1428 CCCCAGGACAGCTGTAGCCCCAGCTCAGCCCGACAGTCCCAACAGAGCAGCCCCGCCAC 1487  
Db 1419 CCCCAGGACAGCTGTAGCCCCAGCTCAGCCCGACAGTCCCAACAGAGCAGCCCCGCCAC 1478  
Qy 1488 CCATCCCCGAGCGCCCGCCACATGAGACACAGCCCTCGCCCTCGCCCGCGCTTTT 1547  
Db 1479 CCATCCCCGAGCGCCCGCCACATGAGACACAGCCCTCGCCCTCGCCCGCGCTTTT 1538  
Qy 1548 CTCTGCCCTTTCTACCGACCATGTGACCCCGACAGCCCTGCCCCCCTGCGCTTCCCGG 1607  
Db 1539 CTCTGCCCTTTCTACCGACCATGTGACCCCGACAGCCCTGCCCCCCTGCGCTTCCCGG 1598  
Qy 1608 GCAGTACTGGGAGCTTCCCTGGGGAGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1667  
Db 1599 GCAGTACTGGGAGCTTCCCTGGGGAGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658  
Qy 1668 AGGCTGGGCGCTCAGTGGACTGCTCTCCACAGCTGCGGCTGACGTGAGAGGCGGAG 1727  
Db 1659 AGGCTGGGCGCTCAGTGGACTGCTCTCCACAGCTGCGGCTGACGTGAGAGGCGGAG 1718  
Qy 1728 GCACGAACTGAGTGGGCGCTGCTTGGGTCTCAGAGTGGGTCTCGGGGGGCTCTGTG 1787  
Db 1719 GCACGAACTGAGTGGGCGCTGCTTGGGTCTCAGAGTGGGTCTCGGGGGGCTCTGTG 1778  
Qy 1788 TTTATTAAGACACCCCTCTGCCAGCTCAGCAATCTTATCACCAGCAACAGCGGAGAC 1847  
Db 1779 TTTATTAAGACACCCCTCTGCCAGCTCAGCAATCTTATCACCAGCAACAGCGGAGAC 1838  
Qy 1848 TTGGCTCCCCCATCTCTCAGAACTCACAAGCATTTGCTCCCCAGCTGGGGAACCTCAACCT 1907  
Db 1839 TTGGCTCCCCCATCTCTCAGAACTCACAAGCATTTGCTCCCCAGCTGGGGAACCTCAACCT 1898  
Qy 1908 CCCCCCTGCCCTGGTGTGACAGAGGGGGGTGGGACAGGGGGGGGTTCCTCCCTCTTAC 1967

Db 1899 CCCCCCTGCTGGTTGGTGACAGAGGGGTGGGACAGGGGGGGTTCGCCCTGTAC 1958  
Qy 1968 ATACCCTGCCATACCAACCCAGGATTAATTCCTCGCTGGTTTGTATTAATTT 2027  
Db 1959 ATACCCTGCCATACCAACCCAGGATTAATTCCTCGCTGGTTTGTATTAATTT 2018  
Qy 2028 TTTTGTGTTGATTTTAAATAAGAAATTTTCATTTTAAGCACAAAAA 2085  
Db 2019 TTTTGTGTTGATTTTAAATAAGAAATTTTCATTTTAAGCACATTTATGAGGAA 2076

## RESULT 9

US-08-592-383-3  
; Sequence 3, Application US/08592383  
; Patent No. 5830760  
; GENERAL INFORMATION:  
; APPLICANT: Tsai, S. and S.J. Collins  
; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acid Recept  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage  
; COMPUTER: IBM PC/386 Compatible  
; OPERATING SYSTEM: MS-DOS 4.01  
; SOFTWARE: Word for Windows 5.01-t  
; CURRENT APPLICATION DATA: US/08/592.383  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/099.242  
; FILING DATE: July 28, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Broderick, Thomas F.  
; REGISTRATION NUMBER: 31,332  
; REFERENCE/DOCKET NUMBER: PHCR-1-7190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2658 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA to mRNA  
; DESCRIPTION: page 4, RAR-alpha403 dominant negative; deleted of 1311-1596 of SEQ.  
; IMMEDIATE SOURCE:  
; LIBRARY: cDNA  
US-08-592-383-3

Query Match 57.5%; Score 1200; DB 2; Length 2658;  
Best Local Similarity 83.5%; Pred. No. 1.2e-233;  
Matches 1501; Conservative 0; Mismatches 15; Indels 282; Gaps 1;

Qy 288 CTCATTGAGACCCAGACGACGATTCGAGAGATAGTGCCACGCCCTCCCTCGCCACC 347  
Db 279 CACCATTTGAGACCCAGACGACGATTCGAGAGATAGTGCCACGCCCTCCCTCGCCACC 338  
Qy 348 CCCTCTACCCCGCATCTACAGCCCTTGTCTGTCTGTCAGGACAGTCTCAGGCTACCA 407  
Db 339 CCCTCTACCCCGCATCTACAGCCCTTGTCTGTCTGTCAGGACAGTCTCAGGCTACCA 398  
Qy 408 CTATGGGGTCAGCGCCTGTGAGGGCTGCAAGGGCTTCTTCGCGCCGACGATCCAGAAGAA 467  
Db 399 CTATGGGGTCAGCGCCTGTGAGGGCTGCAAGGGCTTCTTCGCCCGCAGCATCCAGAAGAA 458

Qy 468 CATGGTGTACACGTGTACCCGGACAGAAAGTGCATCATCAACAAGTGCACCCGGAACCC 527  
Db 459 CATGGTGTACACGTGTACCCGGACAGAAAGTGCATCATCAACAAGTGCACCCGGAACCG 518  
Qy 528 CTGCCAGTACTGCCGACTGCAGAAAGTCTTTGAAGTGGGATGTCCAAAGAGTCTGTGAG 587  
Db 519 CTGCCAGTACTGCCGACTGCAGAAAGTCTTTGAAGTGGGATGTCCAAAGAGTCTGTGAG 578  
Qy 588 AAACGACCCGAAACAGAGAGAGAGAGTGCACCAAGCCGAGTGCCTGTGAGAGCTACAC 647  
Db 579 AAACGACCCGAAACAGAGAGAGAGAGTGCACCAAGCCGAGTGCCTGTGAGAGCTACAC 638  
Qy 648 GCTGACCCGAGGTGGGGAGCTCATTTAGAAAGTGCACCAAGCCGAGTGCCTGTGAGAGCT 707  
Db 639 GCTGACCCGAGGTGGGGAGCTCATTTAGAAAGTGCACCAAGCCGAGTGCCTGTGAGAGCT 698  
Qy 708 CCTGCCCCCTGTCAGCTGGGCAAAATACACTAGCAACACAGCTCAGAAACCTGTGCTC 767  
Db 699 CCTGCCCCCTGTCAGCTGGGCAAAATACACTAGCAACACAGCTCAGAAACCTGTGCTC 758  
Qy 768 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTCCACCAAGTGCATCATTAAGAC 827  
Db 759 TCTGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTCCACCAAGTGCATCATTAAGAC 818  
Qy 828 TGTGGAGTTTCGCCAAGCAGCTGCCGGCTTCACCACTCCACCAAGTGCATCATTAAGAC 887  
Db 819 TGTGGAGTTTCGCCAAGCAGCTGCCGGCTTCACCACTCCACCAAGTGCATCATTAAGAC 878  
Qy 888 CCTCTCAAAGGTGCTGCTGGACATCTCTGCGGATCTGCAAGGGGTACAGGCC 947  
Db 879 CCTCTCAAAGGTGCTGCTGGACATCTCTGCGGATCTGCAAGGGGTACAGGCC 938  
Qy 948 CGAGCAGGACACCATGACCTTCTCGGACGGGCTGACCTGAAACGGGACCCAGATGCACAA 1007  
Db 939 CCAGCAGGACACCATGACCTTCTCGGACGGGCTGACCTGAAACGGGACCCAGATGCACAA 998  
Qy 1008 CGCTGGCTTCGGCCCCCTCACCGACCTGGTCTTTGGCTTCGCAACACAGCTGCTGCCCT 1067  
Db 999 CGCTGGCTTCGGCCCCCTCACCGACCTGGTCTTTGGCTTCGCAACACAGCTGCTGCCCT 1058  
Qy 1068 GGAGATGGATGATGCGGAGACGGGGTGTCTCAGCGGCATCTGCTCATCTCGCGGAGACCG 1127  
Db 1059 GGAGATGGATGATGCGGAGACGGGGTGTCTCAGCGGCATCTGCTCATCTCGCGGAGACCG 1118  
Qy 1128 CCAGGACCTGGAGACGCCGACCGGGTGGACATGCTGTCAGGAGCCGCTGTGAGAGCGCT 1187  
Db 1119 CCAGGACCTGGAGACGCCGACCGGGTGGACATGCTGTCAGGAGCCGCTGTGAGAGCGCT 1178  
Qy 1188 AAAGGTCTACGTGCGGAAGGGAGGCCAGCCCGCCCCACATGTTCCCAAGATGCTAAT 1247  
Db 1179 AAAGGTCTACGTGCGGAAGGGAGGCCAGCCCGCCCCACATGTTCCCAAGATGCTAAT 1238  
Qy 1248 GAAGATTACTGACCTCGGAAGCATCAGCGCAAGGGGGTGTGAGCGGTGTATCAGCTGAA 1307  
Db 1239 GAAGATTACTGACCTCGGAAGCATCAGCGCAAGGGGGTGTGAGCGGTGTATCAGCTGAA 1298  
Qy 1308 GATGAGATCCCGGGTCCATGCGGCTCTCATCCAGGAAATGTTGAGAACTCAGAGGG 1367  
Db 1299 GATGAGATCCGTTAG----- 1312  
Qy 1368 CTTGGACACTGTGAGCGGACAGCCGGGGGTGGGGGGCGGAGCGGGGTGGCCTGCCCCC 1427  
Db 1313 ----- 1312  
Qy 1428 CCGCGCAGCAGCTGTAGCCCCCAGCCTCAGCCCCAGCTCCACAGAGAGCAGCCGCGCAC 1487  
Db 1313 ----- 1312  
Qy 1488 CCACTCCCGTGACCGCCACCGCCACATGACACAGCCCTCGCCCTCCGCCCGCGCTTTT 1547  
Db 1313 ----- 1312  
Qy 1548 CTCTGCCTTTTACCGAACCATGTGACCCCGCACAGCCCTGCCCCCACCTGCCCCCTCCCGG 1607



```

Db      1313      - - - - - CCGG 1316
QY      1608      GCAGTACTGGGACCTTCCTCGGGGACGGGAGGAGGCAGCAGCTCCCTTGACAG 1667
Db      1317      GCAGTACTGGGACCTTCCTCGGGGACGGGAGGAGGCAGCAGCTCCCTTGACAG 1376
QY      1668      AGGCCTGGGCCCTCAGTGACCTGCCTCCACACGCTGGGTGACGTCAGAGGCCGAG 1727
Db      1377      AGGCCTGGGCCCTCAGTGACCTGCCTCCACACGCTGGGTGACGTCAGAGGCCGAG 1436
QY      1728      GCCAGGAATGAGTAGGAGCCCCTGGTCTCGGGTCTCAGATGGGTCTCGGGGCCCTCGTG 1787
Db      1437      GCCAGGAATGAGTAGGAGCCCCTGGTCTCGGGTCTCAGATGGGTCTCGGGGCCCTCGTG 1496
QY      1788      TTCATCAAGACACCCTCTGCCCAGCTCACCATCTTCATCACAGCAAACGCCAGGAC 1847
Db      1497      TTCATCAAGACACCCTCTGCCCAGCTCACCATCTTCATCACAGCAAACGCCAGGAC 1556
QY      1848      TTGGCTCCCCAATCTCAGAATCAAAAGCCATTGCTCCCCAGCTGGGGAACCTCAACCT 1907
Db      1557      TTGGCTCCCCAATCTCAGAATCAAAAGCCATTGCTCCCCAGCTGGGGAACCTCAACCT 1616
QY      1908      CCCCCCTGCCTCGGTGTGTGACAGAGGGGTGGGACAGGGGGGGGGTTCCCCCTGTAC 1967
Db      1617      CCCCCCTGCCTCGGTGTGTGACAGAGGGGTGGGACAGGGGGGGGGTTCCCCCTGTAC 1676
QY      1968      ATACCTGCCATACCAACCCAGGTATTAATTCCTCGCTGGTTTTGTTTTATTTTAATTT 2027
Db      1677      ATACCTGCCATACCAACCCAGGTATTAATTCCTCGCTGGTTTTGTTTTATTTTAATTT 1736
QY      2028      TTTGTTTTGATTTTTTAAATAAGAAATTTTCATTTTAAGACAACAAAAAAA 2085
Db      1737      TTTGTTTTGATTTTTTAAATAAGAAATTTTCATTTTAAGACAATTTATACTGAAGAA 1794

RESULT 10
5260432-1
; Patent No. 5260432
; APPLICANT: TAKAKU, FUMIMARO; ISHIKAWA, TAKASHI; IMAWARI, MICHIO;
; EVANS, RONALD M.; UMESONO, KAZUHIKO
; TITLE OF INVENTION: HUMAN GAMMA RETINOIC ACID RECEPTOR DNA
; NUMBER OF SEQUENCES: 4
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/370,407
; FILING DATE: 22-JUN-1989
; SEQ ID NO:1
; LENGTH: 1576
5260432-1

```

	Query Match	35.9%;	Score 749.2;	DB 6;	Length 1576;
	Best Local Similarity	79.8%;	Pred. No. 1.2e-142;		
	Matches 883;	Conservative 0;	Mismatches 223;	Indels 0;	Gaps 0;
QY	289	TCCATTGAGACCCAGACGAGCGTCTTGAAAGATAGTAGGCCAGCCCTCCCTCGCCACC	348		
Db					
QY	383	TCGGTGGAGACACAGACACCCAGCTCAGAGAGATGTTGCCAGTCTGCCTCGCCCCCT	442		
Db					
QY	349	CCTCTACCCCGCATCTACAAGCCCTTGTCTCTGTCAGGACAAGTCTCTCAGGCTACCAC	408		
Db					
QY	443	CCGCCTCTCTGGGTCTACAAGCCAATGCTTCGTGTGCATGACAAGTCTCTTGGGTATCAC	502		
Db					
QY	409	TATGGGGTCAGCGCCTGTGAGGGCTGCAAGGGCTTCCTCCGCCGAGCATCCAGAAGAAC	468		
Db					
QY	503	TATGGGGTCAGCTCTTCTGAAGGCTGCAAGGGCTCTTTCCGCCAAGCATCCAGAAGAAC	562		
Db					
QY	469	ATGGTGTACACGGTGTACCCGGGACAAGACTGTCATCATCAACAGGTCACCCGGAAACCCC	528		
Db					
QY	563	ATGGTGTACACGGTGTACCCGGCACAAAACCTGTATCATCAACAGGTCACCGAATCGC	622		
Db					
QY	529	TGCCAGTACGTCGCAGCTACAGAGTGTCTTGAAGTGGGCATGTCACAGAGTCTGTGTAGA	588		
Db					
QY	623	TGCCAGTACGTCGCGGTACAGAGTGTCTCAAAGTGGGCATGTCACAGAGTCTGTGCGA	682		
Db					

RESULT 11

```

US-08-392-383-5
; Sequence 5, Application US/08592383
; Patent No. 5830760
;
; GENERAL INFORMATION:
;
; APPLICANT: Tsai, S. and S. J. Collins
; TITLE OF INVENTION: "Hematopoietic (
;
; NUMBER OF SEQUENCES: 10
;
; CORRESPONDENCE ADDRESS:
;
; ADDRESSEE: Christensen, O'Connor,
; STREET: 2800 Pacific First Centre
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98101

```

[illegible]



Qy	945	CCCCGAGGAGACACCATGACCTTCTCGAGCGGCTGACCTGAAACCGGACCGACGATGCA	1004
Db	1134	CCCAAGAACAGACACCATGACTTCTCAGACGGCCTTACCTTAATCGAACTCAGATGCA	1193
Qy	1005	CAACGCTGGCTTGGGCCCCCACCACACCTGCTTTGGCTTCGCCAACACAGCTGCTGCC	1064
Db	1194	CAATGCTGGATTTGGTCTCTGACTGACCTTGTTGCACCTTTGGCAACACAGCTCCCTGCC	1253
Qy	1065	CTTGAGATGGATGATCGGAGACGGGGCTGCTCAGCGCCATCTGCCCTCATCTGGCGAGA	1124
Db	1254	TTTGGAAATGGATGACACAAACAGGCCTTCTCAGTGCATCTCTTAATCTGTGGAGA	1313
Qy	1125	CCGCAGGACCTCGAGACGCGCGGTGCACATGCTCAGAGGACCGCTGCTGGAGGC	1184
Db	1314	CCGCAGGACCTTGAGGAACCGACAAAGTAGATAGCTCAAGAACCATTGCTGGGAAGC	1373
Qy	1185	GCTAAAGGCTACGTGCGGAAGCGGAGGCCACGCGCCACACATGTTCCCAAGATGCT	1244
Db	1374	ACTAAAAATTATATCAGAAAAAGACGCCACCAAGCCTCACATGTTCCAAAGATCTT	1433
Qy	1245	AATGAAGATTACTGACCTCGGAAGCATCAGCCCAAGGGGCTGAGCGGTGATCAGCGT	1304
Db	1434	AATGAAAATCACAGATCTCCGTAGGATCAGTGTCTAAGGTGCAGAGCGTGTGAATTACCTT	1493
Qy	1305	GAAGATGGAGATCCCGGGCTCCATCGCGGCTCTCATCCAGAAATGTTGGAGAACTCAGA	1364
Db	1494	GAAAATGGAAATTCCTGGATCAATGCCACCTCTCATTCAGAAATGATGGAGAAATCTGA	1553
Qy	1365	GGCCTGGACACTCTGAGC	1383
Db	1554	AGGACATGAACCTTGACC	1572

## RESULT 13

```

US-08-896-365-5
Sequence 5, Application US/08896365
Patent No. 5939264
GENERAL INFORMATION:
APPLICANT: Rothschild, Max F.
APPLICANT: Tuggle, Christopher K.
APPLICANT: Messer, Lori A.
APPLICANT: Tun-Ping, Yu
TITLE OF INVENTION: GENES AND GENETIC MARKERS FOR IMPROVED
REPRODUCTIVE TRAITS IN ANIMALS
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Zarley, McKee, Thomte, Voorhees & Sease
STREET: 801 Grand Avenue, Suite 3200
CITY: Des Moines
STATE: Iowa
COUNTRY: USA.
ZIP: 50309
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/896,365
FILING DATE: 18-JUL-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/022,180
FILING DATE: 19-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Nebel, Heidi S.
REGISTRATION NUMBER: 37,719
REFERENCE/DOCKET NUMBER: ISURF 021591
TELECOMMUNICATION INFORMATION:
TELEPHONE: 515-288-3667
TELEFAX: 515-288-1338
INFORMATION FOR SEQ ID NO: 5:

```

```

; SEQUENCE CHARACTERISTICS:
; LENGTH: 558 base pairs
; TYPE: nucleic acid
; STRATEGY: single
; TOPOLOGY: linear
; MOLECULE TYPE: CDNA
; HYPOTHEetical: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: Porcine embryo RARG
; DEVELOPMENTAL STAGE: Embryo
; TISSUE TYPE: Embryo
US-08-896-365-5

Query Match          14.1%; Score 293.4; DB 2; Length 558;
Best Local Similarity 77.2%; Pred. No. le-50;
Matches 429; Conservative 0; Mismatches 121; Indels 6; Gaps

QY 649 CTGACGCGGAGGTGGGGGAGCTCATTGAGAAGGTCCGCAAGCGCACACAGAAAC-CCT 707
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 3 CTTAGCCCCAGTTAGNAGAGCTCATCCACAGGTCAGCAAAAGTCATCAAGAGACGCTC 62
QY 708 CCTGTGCCCTCTGCCAGCTGGCGAAATACACTAGCAACACAGCTCAGAACACGCTGCTC 767
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 63 CTCGCCGCTCTGCCAGCTGGGCAATATACACCAGCAACTCCAGTGTAGACCAACCGGTGCA 122
QY 768 TCTGGACATTGACCTCTGGGA-CAAGTTTCAAGTGAAGTCTCCACCAAGTGCAATTAAGA 826
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 123 GCTGGATCTGGGGCTGTGGGACCAAGTTTCAGTGAAGTGGTCCACCAAGTGCAATTAAGA 182
QY 827 CTGTGGAGTTGCGCAAGCAGCTGCCGGGCTTCACACACCT-CACCATCGCGACACAGATC 885
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 183 TCGCGGAGNTGTCAAGCGGCTGCTCTGTTTTACAGGGGCTCCAGTATTTGCTGACCAAGTC 242
QY 886 ACCCTCTCAAGGCTGCCTCTGGACATCTCTGATCTCGGATCTGCACGCGGTACAG 945
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 243 ACTCTGTCAAGGCTGCCTCTGGACATCTCTGATGCTCGGGNTCTGCACAAAGGTCCACC 302
QY 946 CCC-GAGCAGGACACCATCACTTCTCGACGCGGCTGACCTGAACCGGACCCAGATGCA 1004
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 303 CCCGGCGCAGNTATCATGCACTCTCTGATGGGTGACCTGAAACCGGANCAGATGCA 362
QY 1005 CAACGCTGGCTTC-GGCCCCCTCACGAGCTGGTCTTTGACCTTCGCCAACACAGTGTGC 1063
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 363 CGATGCTGACTTCGGSCCCCTCACAGAGTCTCTGTGGCTTTGCTGGGCAAGTCTCTGC 422
QY 1064 CCTGTGAGATGATATGGCGAGAGCGGGCTCTCTAG-CGGCATCTGCCTCATCTGGCGGA 1122
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 423 CACTGTGAGATGGATGACACAGAGAGAGGGGCTGCTCAGCGCGCATCTGCCTCATCTGGCGGA 482
QY 1123 GACCGCGAGGACTGTGAGCAGCGCGGACCGGCTGACATGCTTGCAGGACCGCTGCTGGAG 1182
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 483 GACCGCATGGACTGTGAGGAACCCAGTAGTAAGTGGCAAGCTGCAAGGCCAATTGCTGGAA 542
QY 1183 GCCTTAAGGCTTACG 1198
      ||| ||| ||| |||
DB 543 GCCTGAGGCTCTATG 558
      ||| ||| ||| |||

```

## RESULT 14

```

RESOLVED 14
US-08-592-383-6
Sequence 6, Application US/08592383
Patent No. 5830760
GENERAL INFORMATION:
APPLICANT: Tsai, S. and S.J. Co
TITLE OF INVENTION: "Hematopole
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christensen, O'Connor
STREET: 2800 Pacific First Center
CITY: Seattle
STATE: Washington
COUNTRY: USA

```

ZIP: 98101  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette-5.25 inch, 1.2mb storage  
COMPUTER: IBM PC/386 Compatible  
OPERATING SYSTEM: MS-DOS 4.01  
SOFTWARE: Word for Windows 5.01-t  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,383  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/099,242  
FILING DATE: July 28, 1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Broderick, Thomas F.  
REGISTRATION NUMBER: 31,332  
REFERENCE/DOCKET NUMBER: FHC-1-7190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 201 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA to mRNA  
DESCRIPTION: page 11, RAR-alpha DNA binding region; positions 364 to 564 of SEQ.  
IMMEDIATE SOURCE:  
LIBRARY: CDNA  
US-08-592-383-6

Query Match 9.6%; Score 199.4; DB 2; Length 201;  
Best Local Similarity 99.5%; Pred. No. 7.5e-32;  
Matches 200; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 373 TGCTTTGCTGTGACGACAGTCTCAGGCTACCACTATGGGTCAGCGCCCTGTGAGGGC 432  
Db 1 TGCTTTGCTGTGACGACAGTCTCAGGCTACCACTATGGGTCAGCGCCCTGTGAGGGC 60  
QY 433 TGAAGGGCTTCTCCCGCCGAGCATCCAGAAACATGGTGTACACGTGTACCGGGAC 492  
Db 61 TGNAGGGCTTCTCCCGCCGAGCATCCAGAAACATGGTGTACACGTGTACCGGGAC 120  
QY 493 AAGAACTGCATCATCAACAAGGTGACCGGAAACCCCTGCCAGTACTGCCGACTGCAGAA 552  
Db 121 AAGAACTGCATCATCAACAAGGTGACCGGAAACCCCTGCCAGTACTGCCGACTGCAGAA 180  
QY 553 TGCTTTGAAGTGGCGATGTC 573  
Db 181 TGCTTTGAAGTGGCGATGTC 201

RESULT 15  
US-08-336-408B-1  
Sequence 1, Application US/08336408B  
Patent No. 5723329  
GENERAL INFORMATION:  
APPLICANT: EVANS, RONALD M.  
APPLICANT: MANGELSDORF, DAVID J.  
TITLE OF INVENTION: RETINOID RECEPTOR COMPOSITIONS AND  
METHODS  
NUMBER OF SEQUENCES: 7  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PRETTY, SCHROEDER, BRUEGGEMANN & CLARK  
STREET: 444 SOUTH FLOWER STREET, SUITE 2000  
CITY: LOS ANGELES  
STATE: CA  
COUNTRY: USA  
ZIP: 90071  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA: US/08/336,408B  
APPLICATION NUMBER: US/08/336,408B  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/933,453  
FILING DATE: 21-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/US91/00399  
FILING DATE: 22-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/478,071  
FILING DATE: 09-FEB-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: REITER, STEPHEN E.  
REGISTRATION NUMBER: 31,192  
REFERENCE/DOCKET NUMBER: P41 9851  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-546-1995  
TELEFAX: 619-546-9392  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1866 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: both  
TOPOLOGY: both  
MOLECULE TYPE: CDNA  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 76...1464  
US-08-336-408B-1  
Query Match 8.8%; Score 183; DB 1; Length 1866;  
Best Local Similarity 50.6%; Pred. No. 3e-28;  
Matches 537; Conservative 0; Mismatches 500; Indels 24; Gaps 3;  
QY 239 CCCAGAGCAGCAGTCTCTGAAGAGATAGTCCCGCAGCCCTCCCTCGCCACCCCTCTACCCC 358  
Db 404 CCTGGGGCTCAATGGCGCTCTCAAGGTCCTCCCGCCACCCCTCAGGAACATGGCTTCCT 463  
QY 359 GCATCTACAAGCTTGTCTGTCTGTCAGGACAAGTCTCAGCTACCACTATATGGGTCA 418  
Db 464 TCACCAAGCAGCATCTGCGCCATCTGCGGGGACCGCTCCTCAGCAAGCAGCTATGGAGTG 523  
QY 419 GCGCTGTGTAGGGCTGCAAGGGCTTCTCCGCGGAGCATCCAGAAAGACATGGTGTACA 478  
Db 524 ACAGCTGCGAGGGTGCAGGGCTTCTTCAAGCGGACGGTGGCGAAGGACCTGACCTACA 583  
QY 479 CGTGTACCGGGGACAGAACTGCATCATCAACAAGGTGACCGGAACCCCTGCCAGTACT 538  
Db 584 CCTGCCGCGACAAAGGACTGCTGATTCAGCAAGCGGACCGGACCGGTCGCGAGTACT 643  
QY 539 GCCGACTGCAGAGTGTCTTGAAGTGGGATGTCCAAGAGTCTGTGAGAAACGACCGAA 598  
Db 644 GCGCTACCAAGAGTGTCTTGGCCATGGGATGAGCGGGAAGCCGTGCAGGAGGAGCG -- 701  
QY 599 ACAAGAAGAAAGAGGTGCCCAAGCCCGAGTGTCTGTAGAGCTACACGCTGACGCGCG 658  
Db 702 -CAGCGTGGCAAGGACCGGAAAGATGAGGTGAGTGCAGCAGCGGCGCAACGAGG 760  
QY 659 AGGTGGGGAGCTCATGTAGAAGGTGCGCAAGCGGACCGGACCGGAAACCTTCCCTGCCCTCT 718  
Db 761 ACATGCGGTGGAGAGGATCTCTGAGGCTGAGCTGGCGGTCGAGCGGCGCAAGACGAGACCT 820  
QY 719 GCCAGTGGGCAAAATACACTAGCAACAACAGCTCAGAAACACGCTGTCTCTGAGACATTG 778  
Db 821 ACCTGGAGGCAAAACATATGGGCTGAACCCCGAGCTC-----GCCGACG 862  
QY 779 ACCTCTGGGACAAAGTTCAGTGAATCTCCCAACCAAGTGCATCATTAAGACTGTGAGTTCG 838  
Db 863 ACCCTGTACCAACATTTGCCAAGCAGCGGACCAACAGCTTTTCACCCCTGTGTGAGTGGG 922

QY 839 CCAAGCAGCTGCCCGGCTTACACACCTCACCATCGCGGACCAAGATACCCCTCTCAAGG 898  
Db 923 CCAAGCGGATCCCAACACTTCTCAGAGTGCCTCGACGACCAAGTCTATCTGTGCGG 982  
QY 899 CTGCTGCTGGACATCTGTGATCTGCGGATCTGACGCGGTACACGCCGAGCAGGACA 958  
Db 983 CAGCTGGAATGAGTGTCTATGCCCTCTTCTCCACCGCTCCATCGCGTGAAGGACG 1042  
QY 959 CCATGACCTTCTCGGAGCGGCTGACCTGAACCGGACCCAGATGCACAGGCTGGCTCG 1018  
Db 1043 GGATCTCTCTGGCCACCGGCTGCACGCTCCACCGGACACGCCGCCACAGCGAGGGTGG 1102  
QY 1019 GCCCTCTACCGACCTGGTCTT---TGCTTTCGCCAACCAAGCTGTCTGCCCTTGAGATGG 1075  
Db 1103 GCCTCATCTTTGACAGGGTGTGACGAGCTTGTGTCCAAGATCGGGACATGCAGATGG 1162  
QY 1076 ATGATCGGAGACGGGCTCTCAGCGCCATCTGCCTCATCTGCGGAGACCGCCAGGACC 1135  
Db 1163 ACAAGAGGAGCTGGGTGCTCGCGCCCATCTGCTCTTTAAACCTGACTCCAAAGGGC 1222  
QY 1136 TGGAGCAGCGGACCGGCTGGACATGCTCAGGAGCGCTGCTGAGGCGCTAAAGGTCT 1195  
Db 1223 TCTCGAACCCGCGGAGGTGAGGCGCTGAGGAGAGGTCTATGCTCTTGGAGCCT 1282  
QY 1196 ACSTGCGGAAGCGGAGGCCAGCGGCCCCACATGTTCCCAAGATGCTAATGAAGATTA 1255  
Db 1283 ACTCAAGCACAAAGTACCCAGAGCAGCGGGAAGTTGCTAAGCTCTTGTCCGCTGC 1342  
QY 1256 CTGACCTGCGAAGCATCAGCGCCAAAGGGGCTGAGCGGTGATCAGCTGAAGATGGAGA 1315  
Db 1343 CGCTCTGCGCTCCATCGGGGCTCAAAATGCTGGAACATCTCTTCTTTTCAAGCTCATCG 1402  
QY 1316 TCCCGGGCTCCATGCGGCTCTCATCCAGAAATGTTGGAG 1356  
Db 1403 GGGACACACCATTGACACCTTCCTTATGGAGATGCTGGAG 1443

Search completed: March 29, 2003, 15:19:14  
Job time : 104 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 29, 2003, 09:44:04 ; Search time 390 seconds  
(without alignments)  
12045.304 Million cell updates/sec

Title: US-09-691-220-1  
Perfect score: 2086  
Sequence: 1 aacagcagcgagcgagggg.....cacaaaaa.....2086

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002:\*

1: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*

2: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*

3: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*

4: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*

5: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*

6: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*

7: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*

8: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*

9: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*

10: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*

11: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*

12: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*

13: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*

14: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*

15: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*

16: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*

17: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*

18: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*

19: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*

20: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*

21: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*

22: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*

23: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*

24: /SIDS2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2086	100.0	2086	24	cdna encoding huma
2	1782.2	85.4	24	AA029338	myl/RAR-alpha fusi
3	1782.2	85.4	3036	19	Human PML/RARalpha
4	1782.2	85.4	3036	20	Human myl/RAR-alpha
5	1782.2	85.4	3511	21	Human PML/RAR-alpha
6	1779	85.3	3511	17	PML-RAR-alpha DNA
7	1778.8	85.3	2907	24	Human CDNA differe
8	1778.8	85.3	2907	24	Lung cancer relate
9	1778.8	85.3	2928	20	Human RAR-alpha cd

10	1777.4	85.2	3511	17	AA033246	PML-RAR-alpha RNA
11	1777.2	85.2	2928	13	AA029338	RAR-alpha gene. S
12	1777.2	85.2	2940	10	AA090124	DNA of clone phRAR
13	1777.2	85.2	2940	16	AA081476	RAR-alpha clone la
14	1707.4	81.9	3052	23	AA583047	DNA encoding novel
15	1540.2	73.8	2930	23	AA583048	DNA encoding novel
16	1500.8	71.9	2073	23	AA585135	DNA encoding novel
17	1500.8	71.9	2457	23	AA583049	DNA encoding novel
18	1499.4	71.9	1944	23	AA583045	DNA encoding novel
19	1499.4	71.9	1944	23	AA585136	DNA encoding novel
20	1207.6	57.9	1692	23	AA583808	DNA encoding novel
21	1200	52.2	1481	17	AA081477	RAR-alpha-403 domi
22	1089.2	52.2	1481	17	AA081477	RAR-alpha RNA sequ
23	1086	52.1	1481	17	AA033260	RAR-alpha RNA sequ
24	749.2	35.9	1576	12	AA010226	Clone pGEM-hRARgam
25	747.6	35.8	1903	12	AA010389	Encodes human Reti
26	747.6	35.8	1903	16	AA05196	Human retinoic aci
27	715.2	34.3	2521	13	AA020542	Mouse retinoic aci
28	715.2	34.3	2740	13	AA020541	Mouse retinoic aci
29	713.6	34.2	1913	16	AA05202	Murine retinoic ac
30	710.4	34.1	1912	12	AA010388	Encodes skin-speci
31	701.2	33.6	1375	24	ABK37492	Human RARbeta2 PCR
32	701.2	33.6	1399	24	ABK37493	Human FLAG RARbeta
33	701.2	33.6	8887	24	ABK37494	Vector genome plas
34	701.2	33.6	9151	24	ABK37495	Vector genome plas
35	674	32.3	2992	16	AAQ79934	Human liver hap cd
36	674	32.3	2992	17	AA06491	Human liver hap cd
37	672.4	32.2	2992	15	AA065572	Human liver hap cd
38	667.8	32.0	2989	10	AA090093	Human Hap (hepatom
39	648.8	31.1	20512	24	AA038339	Genomic DNA encodi
40	591.4	28.4	595	24	AA562827	CDNA sequence #614
41	552	26.5	1351	16	AAQ79935	Human liver RAR-be
42	296.2	14.2	5838	22	AAH98429	Human EST-derived
43	272.8	13.1	301	22	AAF79992	Nucleotide sequenc
44	272.6	13.1	482	23	AA581232	DNA encoding novel
45	271.6	13.0	1641	23	AA586582	DNA encoding novel

ALIGNMENTS

RESULT 1	
AA038338	
ID	AA038338 standard; cdna; 2086 BP.
AC	AA038338;
XX	
XX	
DT	15-AUG-2002 (first entry)
XX	
DE	cdna encoding human nuclear hormone receptor protein.
XX	
KW	Human nuclear hormone receptor; therapeutic agent; gene therapy;
KW	immune response; gene; ss.
XX	
OS	Homo sapiens.
XX	
PH	Key
CDS	Location/Qualifiers
FT	127..1500
FT	/*tag=
FT	/product= "Human nuclear hormone receptor protein"
XX	
XX	WO200231146-A2.
PD	18-APR-2002.
XX	
PF	05-OCT-2001; 2001WO-US31095.
XX	
PR	11-OCT-2000; 2000US-239117P.
PR	19-OCT-2000; 2000US-0691220.
XX	
PA	(PEKE ) PE CORP NY.
XX	
PI	Wei M, Ye J, Yan C, Ketchum KA, Di Francesco V, Beasley EM;

XX WPI: 2002-426282/45.  
 DR P-PSDB; AAO21489.  
 XX  
 PT New human nuclear hormone receptor proteins and nucleic acids, useful  
 PT as models or targets for developing human therapeutic targets, and in  
 PT identifying therapeutic proteins and modulators of nuclear hormone  
 PT receptor expression  
 XX  
 PS Claim 1; Fig 1; 73pp; English.  
 XX  
 CC The invention relates to an isolated peptide of a novel human nuclear  
 CC hormone receptor with a fully defined sequence of 457 amino acids given  
 CC in the specific definition. The novel human nuclear hormone receptor peptides  
 CC and nucleic acids encoding them can be used as models for the development  
 CC of human therapeutic targets, aid in the identification of therapeutic  
 CC proteins, and serve as targets for the development of human therapeutic  
 CC agents that modulate nuclear hormone receptor activity in cells and  
 CC tissues that express the nuclear hormone receptor. The nucleic acids may  
 CC be used as a query sequence to perform searches against sequence  
 CC databases to identify family members or related sequences, as probes or  
 CC primers, to construct recombinant vectors, to identify compounds that  
 CC modulate nuclear hormone receptor nucleic acid expression, in gene  
 CC therapy, and as antisense constructs to control nuclear hormone receptor  
 CC gene expression in cells, tissues or organisms. The polypeptides can be  
 CC used to raise antibodies or to elicit an immune response, as a reagent in  
 CC assays designed to determine protein levels in biological fluids, as  
 CC markers for tissues in which a corresponding protein is expressed, to  
 CC identify a binding partner/ligand to develop a system for the  
 CC identification of inhibitors of the binding reaction, in drug screening  
 CC assays, and to identify compounds that modulate protein activity. This  
 CC polynucleotide sequence represents the cDNA encoding the human nuclear  
 CC hormone receptor protein of the invention.  
 XX  
 SQ Sequence 2086 BP; 437 A; 685 C; 583 G; 381 T; 0 other;

Db	481	TGTCACCGGGACAAAGAACTGCATCATCAACAAAGGTGACCCCGAACCCTTGCCAGTACTGC	540
Qy	541	CGACTGCAGAGTGTCTTTGAAGTGGGCATGTCCAAAGGAGTCTGTGAGAAACACCCGAAAC	600
Db	541	CGACTGCAGAGTGTCTTTGAAGTGGGCATGTCCAAAGGAGTCTGTGAGAAACACCCGAAAC	600
Qy	601	AAGAAGAAGAGGAGGTGCCCCAAGCCCGAGTCTCTGAGAGCTACACGCTGAGCCGGAG	660
Db	601	AAGAAGAAGAGGAGGTGCCCCAAGCCCGAGTCTCTGAGAGCTACACGCTGAGCCGGAG	660
Qy	661	GTGGGGAGCTCATTTGAAAGGTGCGAAAGCGCACACAGAAACCTTCCTGCTCCCTCTGC	720
Db	661	GTGGGGAGCTCATTTGAAAGGTGCGAAAGCGCACACAGAAACCTTCCTGCTCCCTCTGC	720
Qy	721	CAGCTGGGCAATACACTACGAAACACAGCTCAGAACACAGTGTCTCTCTGAGACATTGAC	780
Db	721	CAGCTGGGCAATACACTACGAAACACAGCTCAGAACACAGTGTCTCTCTGAGACATTGAC	780
Qy	781	CTCTGGGCAAGTTCAAGTGAACCTCTCCACCAAGTGCATCATTAAGACTGTGAGTTCGCC	840
Db	781	CTCTGGGCAAGTTCAAGTGAACCTCTCCACCAAGTGCATCATTAAGACTGTGAGTTCGCC	840
Qy	841	AAGCAGCTGCCCCGGCTTCACCAACCTCACCATCGCGGACAGATCACCTCTCTCAAGGCT	900
Db	841	AAGCAGCTGCCCCGGCTTCACCAACCTCACCATCGCGGACAGATCACCTCTCTCAAGGCT	900
Qy	901	GCCTGCTGGACATCTCTGATCTGCGGATCTGCACCGCGGTACACGCCCGGAGGACAC	960
Db	901	GCCTGCTGGACATCTCTGATCTGCGGATCTGCACCGCGGTACACGCCCGGAGGACAC	960
Qy	961	ATGACCTTCTCGGACGGGCTGACCTGAAACCGGACCCAGATGCGACACGCTGGCTTCGCC	1020
Db	961	ATGACCTTCTCGGACGGGCTGACCTGAAACCGGACCCAGATGCGACACGCTGGCTTCGCC	1020
Qy	1021	CCCTCTACCGACCTGGTCTTTGCTTCCTTCGCTCCCAACAGCTGCTGCCCTCGGAGATGAT	1080
Db	1021	CCCTCTACCGACCTGGTCTTTGCTTCCTTCGCTCCCAACAGCTGCTGCCCTCGGAGATGAT	1080
Qy	1081	CGGAGAGCGGGGCTGCTCAGCGCCATCTGCTCATCTCGGAGACCGCCAGACCTGGAG	1140
Db	1081	CGGAGAGCGGGGCTGCTCAGCGCCATCTGCTCATCTCGGAGACCGCCAGACCTGGAG	1140
Qy	1141	CAGCCGACCGGGTGGACATCTGCAGAGCGGCTGCTGGAGGCGCTAAAGTCTACGCTG	1200
Db	1141	CAGCCGACCGGGTGGACATCTGCAGAGCGGCTGCTGGAGGCGCTAAAGTCTACGCTG	1200
Qy	1201	CGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGAAGATTACTGAC	1260
Db	1201	CGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAATGAAGATTACTGAC	1260
Qy	1261	CTGCGAAGCATCAGCGCCAGGGGGCTGAGCGGGTGTACGCTGAAGATGGAGATCCG	1320
Db	1261	CTGCGAAGCATCAGCGCCAGGGGGCTGAGCGGGTGTACGCTGAAGATGGAGATCCG	1320
Qy	1321	GGCTCCATGCGGCTCTCATCCAGGAATGTTGGAGAACTCAGAGGGCTGGACACTCTG	1380
Db	1321	GGCTCCATGCGGCTCTCATCCAGGAATGTTGGAGAACTCAGAGGGCTGGACACTCTG	1380
Qy	1381	AGCGGACAGCGGGGGTGGGGGGCGGAGCGGGGTGGCTGCCCTGCCCGCCGAGGACG	1440
Db	1381	AGCGGACAGCGGGGGTGGGGGGCGGAGCGGGGTGGCTGCCCTGCCCGCCGAGGACG	1440
Qy	1441	TGTAGCCCCAGCCTCAGCCCCAGCTCCCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1500
Db	1441	TGTAGCCCCAGCCTCAGCCCCAGCTCCCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAG	1500
Qy	1501	CCGCCCCAGCCCATGAGACAGCCCTCGGCCCTCCGCCCGGGCTTTTCTCTGCTCTTCTA	1560
Db	1501	CCGCCCCAGCCCATGAGACAGCCCTCGGCCCTCCGCCCGGGCTTTTCTCTGCTCTTCTA	1560
Qy	1561	CCGACCATGTGACCCCGGACAGCCCTGCCCGACCTGCCCTCCGGGCGAGTACTGGGA	1620
Db	1561	CCGACCATGTGACCCCGGACAGCCCTGCCCGACCTGCCCTCCGGGCGAGTACTGGGA	1620



```
Db 1910 AGCAGGACACCATGACCTTCGCGAGCGGCTGACCTGAACCGGAGCCAGATGCACAAGC 1969
QY 1010 CTGCTTGGGCCCCCTACCGACCTGGTCTTTCCTCGCCCAACAGCTGTGCCCCCTGG 1069
Db 1970 CTGCTTGGGCCCCCTACCGACCTGGTCTTTCCTCGCCCAACAGCTGTGCCCCCTGG 2029
QY 1070 AGATGGATGATGCGGAGACGGGCTGCTCAGCGGCATCTGCTCATCTCGGAGACCGCC 1129
Db 2030 AGATGGATGATGCGGAGACGGGCTGCTCAGCGGCATCTGCTCATCTCGGAGACCGCC 2089
QY 1130 AGGACCTGGAGCAGCGCGGCTGGACATGCTGCAGGAGCGGCTGTGAGGCGCTAA 1189
Db 2090 AGGACCTGGAGCAGCGCGGCTGGACATGCTGCAGGAGCGGCTGTGAGGCGCTAA 2149
QY 1190 AGGTCTACGTGCGGAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1249
Db 2150 AGGTCTACGTGCGGAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2209
QY 1250 AGATTACTGACCTCGGAAGCATAGCGCCAAAGGGGGCTGAGCGGTGATCAGCTGAAGA 1309
Db 2210 AGATTACTGACCTCGGAAGCATAGCGCCAAAGGGGGCTGAGCGGTGATCAGCTGAAGA 2269
QY 1310 TGGAGATCCCGGCTCATCGCGGCTCTCATCAGGAAATTTGGAGAACTCAGAGGGCC 1369
Db 2270 TGGAGATCCCGGCTCATCGCGGCTCTCATCAGGAAATTTGGAGAACTCAGAGGGCC 2329
QY 1370 TGGACACTCTGAGGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 1429
Db 2330 TGGACACTCTGAGGGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 2389
QY 1430 CGCCAGGCACTGTAGCGCCAGCTCAGCGCCAGCTCCAAAGAGAGCGGCGGCGGCGGCGG 1489
Db 2390 CGCCAGGCACTGTAGCGCCAGCTCAGCGCCAGCTCCAAAGAGAGCGGCGGCGGCGGCGG 2449
QY 1490 ACTCCCGCTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1549
Db 2450 ACTCCCGCTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2509
QY 1550 CTGCTTCTTACCGACCATGTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1609
Db 2510 CTGCTTCTTACCGACCATGTGACCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 2569
QY 1610 AGTACTGGGACCTTCCCTGGGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1669
Db 2570 AGTACTGGGACCTTCCCTGGGAGCGGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2629
QY 1670 GCCTGGGCGCTCAGTGGACTGCTGCTCCACAGCGCTGAGGCTCAGAGGCGGAGGC 1729
Db 2630 GCCTGGGCGCTCAGTGGACTGCTGCTCCACAGCGCTGAGGCTCAGAGGCGGAGGC 2689
QY 1730 CAGGAAGTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 1789
Db 2690 CAGGAAGTGAAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCT 2749
QY 1790 CATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAAAAGCCAGGACTT 1849
Db 2750 CATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAAAAGCCAGGACTT 2809
QY 1850 GGCTCCCGCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGAACTCAACCTCC 1909
Db 2810 GGCTCCCGCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGAACTCAACCTCC 2869
QY 1910 CCCCTGCTCGGTGGTGACAGAGGGGCTGGACAGGGGCGGGGGTCCCCCTGTACAT 1969
Db 2870 CCCCTGCTCGGTGGTGACAGAGGGGCTGGACAGGGGCGGGGGTCCCCCTGTACAT 2929
QY 1970 ACCCTGCCATACCAACCCAGGATTAATCTCGCTGGTGTGTTTTTAATTTT 2029
Db 2930 ACCCTGCCATACCAACCCAGGATTAATCTCGCTGGTGTGTTTTTAATTTT 2989
QY 2030 TTGTTTTGATTTTTTAAAGAAATTTTCAATTTTAAGCAAAAAA 2076
Db 2990 TTGTTTTGATTTTTTAAAGAAATTTTCAATTTTAAGCAAAAAA 3036
```

## RESULT 3

AAV20474

ID AAV20474 standard; DNA; 3036 BP.

XX AAV20474;

XX 17-JUN-1998 (first entry)

XX Human PML/RARalpha oncogene.

Human; oncogene; proto-oncogene; neoplastic disease; anticancer;  
cancer; antisense oligonucleotide; PML/RARalpha; ds.

XX Homo sapiens.

XX US5734039-A.

XX 31-MAR-1998.

XX 15-SEP-1994; 94US-0306691.

XX 15-SEP-1994; 94US-0306691.

XX (UYJE-) UNIV JEFFERSON THOMAS.

XX Calabretta B, Skorski T;

XX WPI; 1998-229882/20.

Anticancer composition comprising two anti-sense oligo:nucleotide(s)  
- targeting cytoplasmic and nuclear oncogene(s)

Claim 1; Column 145-150; 92pp; English.

The present sequence represents an oncogene from the present invention.  
The present invention describes a composition which comprises two  
antisense oligonucleotides. The first oligonucleotide is specific for a  
cytoplasmic oncogene or proto-oncogene selected from ras, raf, bcr-abl,  
c-fms, c-kit, c-kit, c-kit, c-kit, c-kit, c-kit, c-kit, c-kit, c-kit, c-kit,  
c-yes. The second oligonucleotide is specific for a nuclear oncogene or  
proto-oncogene selected from myc, jun, c-ets, c-fos, c-myc, B-myb,  
c-rel, c-vav, c-ski, c-spl, cyclin D1, PML/RAR alpha, AML1/MTG8,  
E2A/pri and ALL-1/AF-4. The composition is used for treating cancer.  
The combination of antisense oligonucleotides has synergistically  
enhanced ability to inhibit growth of cancer cells.

SQ Sequence 3036 BP; 605 A; 1018 C; 890 G; 523 T; 0 other;

Query Match

Best Local Similarity 85.4%; Score 1782.2; DB 19; Length 3036;

Matches 1784; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 290 CCATTGAGACCCAGAGCAGCAGTCTTGAAGATAGTGCACGAGCTCCCTCGCCACCCC 349

Db 1250 CCATTGAGACCCAGAGCAGCAGTCTTGAAGATAGTGCACGAGCTCCCTCGCCACCCC 1309

QY 350 CTCTACCCCGCATCTACAGCCTTGCTTGTCTCAGGACAAGTCTCAGGCTACCACT 409

Db 1310 CTCTACCCCGCATCTACAGCCTTGCTTGTCTCAGGACAAGTCTCAGGCTACCACT 1369

QY 410 ATGGGCTCAGCGCTGTGAGGCTTGCAGAGGCTTCTTCGCGCGAGCATCCAGAGAAC 469

Db 1370 ATGGGCTCAGCGCTGTGAGGCTTGCAGAGGCTTCTTCGCGCGAGCATCCAGAGAAC 1429

QY 470 TGTGTACAGTGTACACCGGACAAAGTGCATCATCAAGAGTGCACCGGAGACCCCT 529

Db 1430 TGTGTACAGTGTACACCGGACAAAGTGCATCATCAAGAGTGCACCGGAGACCCCT 1489

QY 530 GCCAGTACTGCCAGTGCAGAAAGTGTGTTGAAGTGGGCGATGTCCAAAGAGTCTGTGAGAA 589

Db 1490 GCCAGTACTGCCAGTGCAGAAAGTGTGTTGAAGTGGGCGATGTCCAAAGAGTCTGTGAGAA 1549



```
QY 590 ACGACCGAACAAGAGAGAGAGGTGCCCAAGCCCCAGTGCCTCTGAGAGCTACACGC 649
Db 1550 ACGACCGAACAAGAGAGAGAGGTGCCCAAGCCCCAGTGCCTCTGAGAGCTACACGC 1609
QY 650 TGACCCCGAGGTGGGGAGCTCATTGAGAAGGTGCGCAAGCGCACACGAGGAACCTTCC 709
Db 1610 TGACCCCGAGGTGGGGAGCTCATTGAGAAGGTGCGCAAGCGCACACGAGGAACCTTCC 1669
QY 710 CTGCCCTCTGCGAGCTGGGCAATACATACGAACAACAGCTCAGAACACCTGTCTCTC 769
Db 1670 CTGCCCTCTGCGAGCTGGGCAATACATACGAACAACAGCTCAGAACACCTGTCTCTC 1729
QY 770 TGGACATTGACCTCTGGGACAGTTTCAGTGAAGTCTCCACCAAGTGCATTAAGACTG 829
Db 1730 TGGACATTGACCTCTGGGACAGTTTCAGTGAAGTCTCCACCAAGTGCATTAAGACTG 1789
QY 830 TGGAGTTGCCCAAGCAGCTGCCCGCTTCACCAACCTTCACCATCGCGACCAAGTCAACC 889
Db 1790 TGGAGTTGCCCAAGCAGCTGCCCGCTTCACCAACCTTCACCATCGCGACCAAGTCAACC 1849
QY 890 TCCTCAAGGCTGCCCTGGACATCTCTGATCCTCGGATCTGCACGCGGTACACGCCCG 949
Db 1850 TCCTCAAGGCTGCCCTGGACATCTCTGATCCTCGGATCTGCACGCGGTACACGCCCG 1909
QY 950 AGCAGGACACCATGACCTTCTCGGACGGGCTGACCCCTGAACCGGACCCAGATGCAACG 1009
Db 1910 AGCAGGACACCATGACCTTCTCGGACGGGCTGACCCCTGAACCGGACCCAGATGCAACG 1969
QY 1010 CTGGCTCGGCCCCCTACCGACCTTGTCTTGGCTTGGCCCTGCGCAACCAAGCTGCTGCCCTGG 1069
Db 1970 CTGGCTCGGCCCCCTACCGACCTTGTCTTGGCTTGGCCCTGCGCAACCAAGCTGCTGCCCTGG 2029
QY 1070 AGATGGATGATCGGAGACGGGGCTGCTCAGCGCATCTGCTCATCTCGCGGAGACCGCC 1129
Db 2030 AGATGGATGATCGGAGACGGGGCTGCTCAGCGCATCTGCTCATCTCGCGGAGACCGCC 2089
QY 1130 AGGACCTGGAGACGGCGGCTGGACATGCTCGAGAGCGGCTGCTGGAGGGCGCTAA 1189
Db 2090 AGGACCTGGAGACGGCGGCTGGACATGCTCGAGAGCGGCTGCTGGAGGGCGCTAA 2149
QY 1190 AGGTCTAGTGGGAAGCGGAGGCCCGCCCGCCACATGTTCCCAAGATGCTAAATGA 1249
Db 2150 AGGTCTAGTGGGAAGCGGAGGCCCGCCCGCCACATGTTCCCAAGATGCTAAATGA 2209
QY 1250 AGATTACTGACCTGGAAGCATCAGCGCAAGGGGGCTGAGCGGTGATCAGCTGAAGA 1309
Db 2210 AGATTACTGACCTGGAAGCATCAGCGCAAGGGGGCTGAGCGGTGATCAGCTGAAGA 2269
QY 1310 TGGAGATCCCGGGTCCATCGCGCTCTCATCCAGGAATGTTGGAGAACTCAGAGGGCC 1369
Db 2270 TGGAGATCCCGGGTCCATCGCGCTCTCATCCAGGAATGTTGGAGAACTCAGAGGGCC 2329
QY 1370 TGGACATCTGAGCGGACAGCGGGGGGTGGGGGCGGGAGCGGGGTGCGCTGCCCTGCC 1429
Db 2330 TGGACATCTGAGCGGACAGCGGGGGGTGGGGGCGGGAGCGGGGTGCGCTGCCCTGCC 2389
QY 1430 CGCCAGGAGCTGTAGCCCGCCAGCTCAGCGCCAGCTCCAAACAAGAGCGCCCGCCACCC 1489
Db 2390 CGCCAGGAGCTGTAGCCCGCCAGCTCAGCGCCAGCTCCAAACAAGAGCGCCCGCCACCC 2449
QY 1490 ACTCCCGGTGACCGCCACGACATGAGACAGCCCTCGCCCTCGCGCCCGCGCTTTCT 1549
Db 2450 ACTCCCGGTGACCGCCACGACATGAGACAGCCCTCGCCCTCGCGCCCGCGCTTTCT 2509
QY 1550 CTGCCCTTCTACGACCATGTGACCCCGACAGCCCTGCCCCCAGCTGCCCTCCCGGGC 1609
Db 2510 CTGCCCTTCTACGACCATGTGACCCCGACAGCCCTGCCCCCAGCTGCCCTCCCGGGC 2569
QY 1610 AGTACTGGGGACCTTCCCTGGGGACGGGGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1669
Db 2570 AGTACTGGGGACCTTCCCTGGGGACGGGGAGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2629
```

```
QY 1670 GCCTGGGCCCTCAGTGGACTGCTGCTCCACACAGCTGGGCTGACGTCAGAGGCGGAGGC 1729
Db 2630 GCCTGGGCCCTCAGTGGACTGCTGCTCCACACAGCTGGGCTGACGTCAGAGGCGGAGGC 2689
QY 1730 CAGGAACCTGAGTCAGGCGCCCTGGTCTCGGTCTCAGGATGGGTCTCGGGGCTCGTGT 1789
Db 2690 CAGGAACCTGAGTCAGGCGCCCTGGTCTCGGTCTCAGGATGGGTCTCGGGGCTCGTGT 2749
QY 1790 CATCAAGACACCCCTCTGCCCCAGCTCACCACATCTTCATCACCAGCAAAACCCAGGACTT 1849
Db 2750 CATCAAGACACCCCTCTGCCCCAGCTCACCACATCTTCATCACCAGCAAAACCCAGGACTT 2809
QY 1850 GGCTCCCCATCTCAGAACTCACAAGCCATTCCTCCCCAGCTGGGAACCTCAACCTCC 1909
Db 2810 GGCTCCCCATCTCAGAACTCACAAGCCATTCCTCCCCAGCTGGGAACCTCAACCTCC 2869
QY 1910 CCCTCCCTCGGTGGTGCAGAGGGGGTGGACAGGGCGGGGGTTCGCCCTGTACAT 1969
Db 2870 CCCTCCCTCGGTGGTGCAGAGGGGGTGGACAGGGCGGGGGTTCGCCCTGTACAT 2929
QY 1970 ACCCTGCCATACCAACCCAGGTATTAAATCTCGTGGTGTGTTTTTATTTTATTTT 2029
Db 2930 ACCCTGCCATACCAACCCAGGTATTAAATCTCGTGGTGTGTTTTTATTTTATTTT 2989
QY 2030 TTGTTTTGATTTTTTAAAGAATTTTCATTTTAAGCACAAAAA 2076
Db 2990 TTGTTTTGATTTTTTAAAGAATTTTCATTTTAAGCAAAAAA 3036

RESULT 4
AAV64990
ID AAV64990 standard; cDNA; 3036 BP.
XX
AC AAV64990;
XX
AC AAV64990;
DT 05-FEB-1999 (first entry)
XX
DE Human myl/RAR-alpha fusion protein cDNA.
XX
KW Fusion protein; myl; retinoic acid receptor-alpha; RAR; human;
KW acute promyelocytic leukaemia; APL; t(15;17); translocation;
KW treatment; all-trans retinoic acid; ss.
XX
OS Homo sapiens.
XX
FH Key
FT CDS
FT Location/Qualifiers
FT 67..2460
FT /*tag= a
FT /product= "myl/RAR-alpha fusion protein"
XX
PN US5843642-A.
XX
PD 01-DEC-1998.
XX
PF 21-JUL-1993; 93US-0095728.
XX
PR 21-JUL-1993; 93US-0095728.
PR 22-MAR-1991; 91US-0673838.
PR 22-MAR-1991; 91US-0675084.
XX
PA (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI Dmitrovsky E, Frankel S, Miller WH, Warrell RP;
XX
XX WPI; 1999-044563/04.
DR P-PSDB; AAW81963.
XX
PT Diagnosis of acute promyelocytic leukaemia - by detecting nucleic
PT acid encoding abnormal retinoic acid receptor-alpha
XX
PS Disclosure; Fig 4A-G; 38pp; English.
XX
```

CC This sequence encodes a human myl/RAR-alpha (retinoic acid receptor  
CC alpha) fusion protein which is used in a method for identifying a  
CC subject with acute promyelocytic leukaemia (APL) resulting from a  
CC t(15;17) translocation who will respond to treatment with all-trans  
CC retinoic acid. The protein can also be used to identify a subject  
CC with indications of APL who will not respond to treatment with all-trans  
CC retinoic acid.

XX  
SQ Sequence 3036 BP; 605 A; 1018 C; 890 G; 523 T; 0 other;

Query Match 85.48; Score 1782.2; DB 20; Length 3036;  
Best Local Similarity 99.8%; Pred. No. 0;  
Matches 1784; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

```
QY 290 CCATTGAGACCCAGAGCAGCTTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACCCC 349
DB 1250 CCATTGAGACCCAGAGCAGCTTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACCCC 1309
QY 350 CTCTACCCCGCATCTACAAGCCTTGTCTGTCTGAGCAAGTCTCAGGCTACCACT 409
DB 1310 CTCTACCCCGCATCTACAAGCCTTGTCTGTCTGAGCAAGTCTCAGGCTACCACT 1369
QY 410 ATGGGGTTCAGGGCTGTGAGGGCTGTCAAGGGCTTCTCCGCGCAGCATCCAGAAGAACA 469
DB 1370 ATGGGGTTCAGGGCTGTGAGGGCTGTCAAGGGCTTCTCCGCGCAGCATCCAGAAGAACA 1429
QY 470 TGTGTACAGTGTACCCGGGACAGACTGCATCATCAACAAGTGTACCCGGGAACCCCT 529
DB 1430 TGTGTACAGTGTACCCGGGACAGACTGCATCATCAACAAGTGTACCCGGGAACCCCT 1489
QY 530 GCCAGTACTGCCAGTCTCAGAAAGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAGAA 589
DB 1490 GCCAGTACTGCCAGTCTCAGAAAGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAGAA 1549
QY 590 ACGACCGAAACAAGAAAGAGAGTGCCCAAGCCCGAGTGTCTGAGAGCTACACGC 649
DB 1550 ACGACCGAAACAAGAAAGAGAGTGCCCAAGCCCGAGTGTCTGAGAGCTACACGC 1609
QY 650 TGACGGCGGAGGTGGGGAGCTATTGAGAGGTGGCCAAAGCCGACCAAGAAACCTTCC 709
DB 1610 TGACGGCGGAGGTGGGGAGCTATTGAGAGGTGGCCAAAGCCGACCAAGAAACCTTCC 1669
QY 710 CTGCCCTCTGCCAGTGGGCAATATACACTACGAACAACAGCTCAGAAACAGTCTCTCTC 769
DB 1670 CTGCCCTCTGCCAGTGGGCAATATACACTACGAACAACAGCTCAGAAACAGTCTCTCTC 1729
QY 770 TGGACATTGACCTCTGGGACAAGTTCAGTGAACCTCTCCACCAAGTGTGATTAAGACTG 829
DB 1730 TGGACATTGACCTCTGGGACAAGTTCAGTGAACCTCTCCACCAAGTGTGATTAAGACTG 1789
QY 830 TGGAGTTCGCCAAGCAGCTGCCGGCTTACACCCCTCACCATGCGCGGACAGATACCC 889
DB 1790 TGGAGTTCGCCAAGCAGCTGCCGGCTTACACCCCTCACCATGCGCGGACAGATACCC 1849
QY 890 TCCTCAAGGTGTGCTGTGCATCTGATCTGCGGATCTGCACGCGGTACACGCCCG 949
DB 1850 TCCTCAAGGTGTGCTGTGCATCTGATCTGCGGATCTGCACGCGGTACACGCCCG 1909
QY 950 AGCAGACACCATGACCTTCTCGGAGGGGTGACCTGAAACCGGACCCAGATGCACAAG 1009
DB 1910 AGCAGACACCATGACCTTCTCGGAGGGGTGACCTGAAACCGGACCCAGATGCACAAG 1969
QY 1010 CTGGCTTCGGCCCTTCACCGACCTGGTCTTTCGCTTCGCGCAACCACTGCTGCCCTGG 1069
DB 1970 CTGGCTTCGGCCCTTCACCGACCTGGTCTTTCGCTTCGCGCAACCACTGCTGCCCTGG 2029
QY 1070 AGATGATGATCGGAGACGGGGTCTCTAGCGCCATCTGCCCTCATCTCGGGAGACGCC 1129
DB 2030 AGATGATGATCGGAGACGGGGTCTCTAGCGCCATCTGCCCTCATCTCGGGAGACGCC 2089
QY 1130 AGGACTTGGAGCAGCGGACCGGGTGGACATGCTGAGGAGCGGCTGCTGGAGCGGTAA 1189
DB 2090 AGGACTTGGAGCAGCGGACCGGGTGGACATGCTGAGGAGCGGCTGCTGGAGCGGTAA 2149
```

```
QY 1190 AGGTCTAGCTGCGGAGCGAGCGCCAGCGCCACATGTTCCCAAGATGCTAATGA 1249
DB 2150 AGGTCTAGCTGCGGAGCGAGCGCCAGCGCCACATGTTCCCAAGATGCTAATGA 2209
QY 1250 AGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGAGCGGGTGTATCAGCTGAAGA 1309
DB 2210 AGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGAGCGGGTGTATCAGCTGAAGA 2269
QY 1310 TGGAGATCCCGGGCTCCATGCGCCCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGGCC 1369
DB 2270 TGGAGATCCCGGGCTCCATGCGCCCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGGCC 2329
QY 1370 TGGACACTCTGAGCGGACAGCGCGGGGGTGGGGGGGACGCGGGTGGCTGCCCGCCC 1429
DB 2330 TGGACACTCTGAGCGGACAGCGCGGGGGTGGGGGGGACGCGGGTGGCTGCCCGCCC 2389
QY 1430 CGCCAGGAGCTGTAGCCCGCAGCCTCAGCCCCAGCTCCCAACAGAGCAGCGCGGCCACCC 1489
DB 2390 CGCCAGGAGCTGTAGCGCCCGCAGCCTCAGCCCCAGCTCCCAACAGAGCAGCGCGGCCACCC 2449
QY 1490 ACTCCCGCTGACCGCCCGCAGCCTGACACACAGCCCTCGCCCTCGCCCGCGCTTTTCT 1549
DB 2450 ACTCCCGCTGACCGCCCGCAGCCTGACACACAGCCTCGCCCTCGCCCGCGCTTTTCT 2509
QY 1550 CTGCCCTTCTACCGGACATGTACCCCGGACAGCGCCCTCGCCCTCGCCCGCGGCG 1609
DB 2510 CTGCCCTTCTACCGGACATGTACCCCGGACAGCGCCCTCGCCCTCGCCCGCGGCG 2569
QY 1610 AGTACTGGGGACCTTCCCTGGGGGACGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1669
DB 2570 AGTACTGGGGACCTTCCCTGGGGGACGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 2629
QY 1670 GCCTGGGGCCCTCAGTGGAGTGCCTCTCCACAGCCTTGGGCTGAGCTCAGAGGCCGAGGC 1729
DB 2630 GCCTGGGGCCCTCAGTGGAGTGCCTCTCCACAGCCTTGGGCTGAGCTCAGAGGCCGAGGC 2689
QY 1730 CAGGAAGTGTAGAGGCCCTGGTCTGGTCTCAGGATGGGTCTGGGGGCTCTGTGT 1789
DB 2690 CAGGAAGTGTAGAGGCCCTGGTCTGGTCTCAGGATGGGTCTGGGGGCTCTGTGT 2749
QY 1790 CATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAACGCGGAGCTT 1849
DB 2750 CATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAACGCGGAGCTT 2809
QY 1850 GGCTCCCGCCATCTCAGAACTCAGAGCCATTGCTCCCGAGCTGGGGAACCTCAACCTCC 1909
DB 2810 GGCTCCCGCCATCTCAGAACTCAGAGCCATTGCTCCCGAGCTGGGGAACCTCAACCTCC 2869
QY 1910 CCCCTGCTCGGTTGGTGACAGAGGGGTGGGACAGGGGGGGGGTTCGCCCTCTACAT 1969
DB 2870 CCCCTGCTCGGTTGGTGACAGAGGGGTGGGACAGGGGGGGGGTTCGCCCTCTACAT 2929
QY 1970 ACCCTGCATACCAACCCAGGATTAATCTCGCTGTTTGTGTTTATTTATTTT 2029
DB 2930 ACCCTGCATACCAACCCAGGATTAATCTCGCTGTTTGTGTTTATTTATTTT 2989
QY 2030 TTGTTTGTATTTTATTAAGAAATTTTCAATTTTAAGCAAAAAA 2076
DB 2990 TTGTTTGTATTTTATTAAGAAATTTTCAATTTTAAGCAAAAAA 3036
```

RESULT 5  
AAA38656  
ID AAA38656 standard; DNA; 3511 BP.  
XX  
AC AAA38656;  
XX  
DT 11-SEP-2000 (first entry)  
XX Human PML/alpha-RAR fusion gene, SEQ ID NO:13.  
XX External guide sequence; EGS; RNase P; human; ribozyme;  
KW

KW	catalytic RNA; antiviral; hepatitis B virus; HBV; anticancer; APL;
KW	acute promyelocytic leukaemia; retinoic acid receptor-alpha;
KW	RAR-alpha; PML; fusion gene; ss.
OS	Homo sapiens.
XX	
PN	US6057153-A.
XX	
PD	02-MAY-2000.
XX	
PF	14-JUL-1997; 97US-0892747.
XX	
PR	13-JAN-1995; 95US-0372556.
PR	19-JAN-1996; 96WO-US00513.
XX	(UYVA ) UNIV YALE.
PA	
PI	Pace U, Goldberg AR, Werner M, George ST, Ma M;
XX	WPI; 2000-399010/34.
XX	
PT	Novel modified external guide sequences for eukaryotic RNase P for
PT	cleaving hepatitis viral RNA comprises RNase P targeting sequence and
PT	recognition sequence complementary to targeted sequence in target RNA
PT	.
XX	
PS	Example 1; Column 45-48; 76pp; English.
XX	
CC	The invention relates to modified human RNase P external guide sequences
CC	(EGSS). The EGS is a ribozyme which forms the catalytic domain of the
CC	ribonucleoprotein RNase P, which is involved in the processing of tRNA.
CC	The EGS comprises a cleavage targeting sequence and a recognition
CC	sequence which is complementary to the target RNA sequence. The
CC	recognition sequence comprises an A and a D recognition arm which
CC	correspond to the aminoacyl acceptor stem and the D stem of tRNA. The A
CC	recognition arm is located at the 3' end of the EGS and the D at the 5'
CC	end. The cleavage targeting sequence forms a structure corresponding to
CC	the T stem and T loop of precursor tRNA. The modified EGSs of the
CC	invention comprise at least one modified nucleotide or an unmodified
CC	deoxyribonucleotide, and the 3' end of the molecule is preferably
CC	modified. The A and D arms, and the T stem and loop can consist of
CC	ribonucleotides with 5'-phosphate or 5'-phosphorothioate linkages;
CC	2'-modified ribonucleotides with 5'-phosphate or 5'-phosphorothioate
CC	linkages; or deoxyribonucleotides with 5'-phosphate or 5'-
CC	phosphorothioate linkages. EGSs of the invention are used for cleaving
CC	a target RNA molecule. The invention discloses EGSs which are targeted
CC	to hepatitis B virus (HBV) RNA sequences, particularly hepatitis B
CC	surface antigen (HBSAg) RNA, and EGSs targeted at PML/retinoic acid
CC	receptor alpha (RAR-alpha) fusion gene mRNA which is found in the
CC	majority of acute promyelocytic leukaemia (APL) cases. Sequences
CC	AAA38652- AAA38655 represent active and control PML/RAR-alpha fusion gene
CC	RNA- targeted EGSs used in an exemplification of the invention. The
CC	target RNA (AAA38651) comprises the transcript of the junction between
CC	the PML and the RAR-alpha gene fragments of the fusion gene. The fusion
CC	gene is generated by a balanced reciprocal chromosomal translocation
CC	between the long arms of chromosome 15 and 17, and the fusion protein
CC	encoded is a dominant negative regulator of myeloid differentiation.
XX	
SQ	Sequence 3511 BP; 715 A; 1168 C; 1038 G; 589 T; 1 other;
	Query Match 85.4%; Score 1782.2; DB 21; Length 3511;
	Best Local Similarity 99.8%; Pred. No. 0;
	Matches 1784; Conservative 0; Mismatches 3; Indels 0; Gaps 0
QY	290 CCATTGAGACCCAGAGCAGTTCCTGAAGAGATAGTGCCCGAGCCCTCCCTGCCACCCC 349
	1725 CCATTGAGACCCAGAGCAGTTCCTGAAGAGATAGTGCCCGAGCCCTCCCTGCCACCCC 1784
Dd	
QY	350 CTCTACCCCGCATCTACAAGCCTTCCTTTGTCTGTGCAGCACAGTCCTCAGGCTACCCT 409
Dd	1785 CTCFACCCCGCATCTACAAGCCTTCCTTTGTCTGTGCAGCACAGTCCTCAGGCTACCCT 1844
QY	410 ATGGGGTCAGCGCCTGTGAGGGCTCCAAAGGGCTTCTTCGCGCGCAGCATCTCCAGAAGCA 459

```
Db 2925 ACTCCCGTGACCGCCACACCCACATGACACAGCCCTCGCCCTCCCGCGGCTTTTCT 2984
QY 1550 CTGCTTTCTACGACCATGTGACCCCGACACAGCCCTGCCCCACCTGCCCTCCCGGGC 1609
Db 2985 CTGCTTTCTACGACCATGTGACCCCGACACAGCCCTGCCCCACCTGCCCTCCCGGGC 3044
QY 1610 AGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1669
Db 3045 AGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 3104
QY 1670 GCTGGGCGCTCAGTGGAGCTGCTGCTCCACAGCTGGGCTGACGCTGACGCTGACGCTG 1729
Db 3105 GCTGGGCGCTCAGTGGAGCTGCTGCTCCACAGCTGGGCTGACGCTGACGCTGACGCTG 3164
QY 1730 CAGGAAGTACTGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1789
Db 3165 CAGGAAGTACTGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3224
QY 1790 CATCAAGACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1849
Db 3225 CATCAAGACACCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3284
QY 1850 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1909
Db 3285 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3344
QY 1910 CCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1969
Db 3345 CCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3404
QY 1970 ACCCTGCCATACAAACCCAGGATTAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2029
Db 3405 ACCCTGCCATACAAACCCAGGATTAATTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3464
QY 2030 TTGTTTGTGTTTATTAAGAAATTTTCATTTTAAAGCAAAAAA 2076
Db 3465 TTGTTTGTGTTTATTAAGAAATTTTCATTTTAAAGCAAAAAA 3511

RESULT 6
AAT33259
ID AAT33259 standard; cDNA; 3511 BP.
XX AC AAT33259;
XX DT 09-OCT-1996 (first entry)
XX DE PML-RAR-alpha DNA sequence.
XX KW Hammerhead ribozyme; acute promyelocytic leukaemia; APL;
XX KW lymphoma; therapy; PML-RAR-alpha; retinoic acid receptor;
XX KW external guide sequence; EGS; antisense; ss.
XX OS Homo sapiens.
XX FH Key
XX misc_feature 1..1724 Location/Qualifiers
FT FT /*tag= a
FT FT /note= "PML DNA"
FT FT misc_difference 1477
FT FT /*tag= b
FT FT /note= "base n at position 1477 is not identified
FT FT /note= "in the specification"
FT FT misc_feature 1725..3511
FT FT /*tag= c
FT FT /note= "RAR-alpha DNA"
XX PN WO9618733-A2.
XX PD 20-JUN-1996.
XX PF 14-DEC-1995; 95WO-US16451.
XX XX
```

```
PR 14-DEC-1994; 94US-0354956.
XX (INNO-) INNOVIR LAB INC.
XX George ST, Goldberg AR, Pace U;
XX WPI; 1996-300650/30.
XX RNA construct(s) including ribozyme(s) and antisense
XX oligo:nucleotide(s) - for the inactivation of RNA associated with,
XX e.g. promyelocytic leukaemia or follicular lymphoma
XX Example 1; Page 50-52; 81pp; English.
XX PML-RAR-alpha DNA (AAT33259) is associated with acute promyelocytic
XX leukemia (APL). APL is characterised by a balanced, reciprocal
XX translocation between the long arms of chromosomes 15 and 17,
XX resulting in a fusion of the retinoic acid receptor gene (RAR-alpha)
XX and a gene for a putative transcription factor, PML. Methods for
XX treating APL involve the use of ribozymes, external guide sequences
XX and antisense oligonucleotides (see also AAT33245 and AAT33251) that
XX specifically cleave the PML-RAR-alpha fusion mRNA (see also AAT33251)
XX but not wild-type RAR-alpha mRNA. Plasmids carrying portions of the
XX PML-RAR-alpha and RAR-alpha genes (see also AAT33260) allow synthesis
XX of shortened versions of APL mRNA in vitro, facilitating testing and
XX screening processes.
XX SQ Sequence 3511 BP; 715 A; 1168 C; 1038 G; 589 T; 1 other;
Query Match 85.3%; Score 1779; DB 17; Length 3511;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 1782; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 290 CCATTGAGACCCAGACGAGCTTCTGAAGAGATAGTGCCAGCCCTCCCTCGCCACCCC 349
Db 1725 CCATTGAGACCCAGACGAGCTTCTGAAGAGATAGTGCCAGCCCTCCCTCGCCACCCC 1784
QY 350 CTCCTACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAAGTCTCAGGTACCAC 409
Db 1785 CTCCTACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAAGTCTCAGGTACCAC 1844
QY 410 ATGGGGTCAGCGCCTGTGAGGGCTGCAAGGGCTTCTCCCGCCGAGCATCCAGAAACA 469
Db 1845 ATGGGGTCAGCGCCTGTGAGGGCTGCAAGGGCTTCTCCCGCCGAGCATCCAGAAACA 1904
QY 470 TGGTGTACAGTGTCAACCGGGACAAAGTGTGATCATCAACAAGGTGACCCGAAACCCCT 529
Db 1905 TGGTGTACAGTGTCAACCGGGACAAAGTGTGATCATCAACAAGGTGACCCGAAACCCCT 1964
QY 530 GCCAGTACTGCCAGCTGCAAGAGTGTCTTGAAGTGGGAGTGTCCAAGGAGTCTGTGAGAA 589
Db 1965 GCCAGTACTGCCAGCTGCAAGAGTGTCTTGAAGTGGGAGTGTCCAAGGAGTCTGTGAGAA 2024
QY 590 ACGACCGAAACAAGAGAGAGGAGTCCCAAGCCCGAGTGTCTGTGAGAGTACACGC 649
Db 2025 ACGACCGAAACAAGAGAGAGGAGTCCCAAGCCCGAGTGTCTGTGAGAGTACACGC 2084
QY 650 TGACCCCGGAGTGGGGAGCTCATTTGAGAAGGTGCGCAAGCGCACCAAGAAACCTTCC 709
Db 2085 TGACCCCGGAGTGGGGAGCTCATTTGAGAAGGTGCGCAAGCGCACCAAGAAACCTTCC 2144
QY 710 CTGCCCCCTGCCAGCTGGGCAATATCACTACGAACAACAGCTCAGAACACGCTGTCTCTC 769
Db 2145 CTGCCCCCTGCCAGCTGGGCAATATCACTACGAACAACAGCTCAGAACACGCTGTCTCTC 2204
QY 770 TGGACATTGACCTCTGGGACAAAGTTCACTGTAAGTCTCCACCAAGTGCATCTTAAGACTG 829
Db 2205 TGGACATTGACCTCTGGGACAAAGTTCACTGTAAGTCTCCACCAAGTGCATCTTAAGACTG 2264
QY 830 TGGAGTTCGCCAAGCAGCTGCCCGGCTTCAACCACCTCACCATTGCCCGACCATCACCC 889
Db 2265 TGGAGTTCGCCAAGCAGCTGCCCGGCTTCAACCACCTCACCATTGCCCGACCATCACCC 2324
```

QY 890 TCCTCAAGGCTGCTGCTGACATCTGATCTCTGGGATCTGCACGCGGTACACGCCG 949  
DB 2325 TCCTCAAGGCTGCTGCTGACATCTGATCTCTGGGATCTGCACGCGGTACACGCCG 2384  
QY 950 AGCAGACACCATGACCTTCTCGGAGGGGTGACCTTGAACCGGACCCAGATGCACACG 1009  
DB 2385 AGCAGACACCATGACCTTCTCGGAGGGGTGACCTTGAACCGGACCCAGATGCACACG 2444  
QY 1010 CTGGCTTCGGCCCTTACCGACCTGCTCTTGGCTTCGCTCGCAACCAAGCTGCTGCCCTGG 1069  
DB 2445 CTGGCTTCGGCCCTTACCGACCTGCTCTTGGCTTCGCTCGCAACCAAGCTGCTGCCCTGG 2504  
QY 1070 AGATGATGATCGGAGACGGGCTGCTCAGCGCCATCTGCCCTCATCTGCGGAGACCGCC 1129  
DB 2505 AGATGATGATCGGAGACGGGCTGCTCAGCGCCATCTGCCCTCATCTGCGGAGACCGCC 2564  
QY 1130 AGACCTGAGACGCGGAGCGGGTGGACATGCTGCAGGAGCGGCTGCTGGAGCGGCTAA 1189  
DB 2565 AGACCTGAGACGCGGAGCGGGTGGACATGCTGCAGGAGCGGCTGCTGGAGCGGCTAA 2624  
QY 1190 AGTCTACCTGCGAAGCGGAGCGCCAGCGCCGCCCATCTGTTCCCAAGATGCTAATGA 1249  
DB 2625 AGTCTACCTGCGAAGCGGAGCGCCAGCGCCGCCCATCTGTTCCCAAGATGCTAATGA 2684  
QY 1250 AGATTACTGACCTTGCAGACATCAGCGCCAAAGGGGCTGAGCGGGTGCATCAGCTGAAGA 1309  
DB 2685 AGATTACTGACCTTGCAGACATCAGCGCCAAAGGGGCTGAGCGGGTGCATCAGCTGAAGA 2744  
QY 1310 TGGAGATCCGGGCTCCATCGCCCTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCC 1369  
DB 2745 TGGAGATCCGGGCTCCATCGCCCTCTCATCCAGAAATGTTGGAGAACTCAGAGGGCC 2804  
QY 1370 TGGACACTTGAGCGGACAGCGGGGGTGGGGGGCGGACGGGGTGGCTGCCCTCCGCC 1429  
DB 2805 TGGACACTTGAGCGGACAGCGGGGGTGGGGGGCGGACGGGGTGGCTGCCCTCCGCC 2864  
QY 1430 CGCCAGGAGCTGTAGCCCGACCTCAGCGCCAGCTCCACAGAAAGCGCGGCGCACCC 1489  
DB 2865 CGCCAGGAGCTGTAGCCCGACCTCAGCGCCAGCTCCACAGAAAGCGCGGCGCACCC 2924  
QY 1490 ACTCCCGTGGCCCGCCAGCCACATGAGACAGCCCTCGCCCTCGCCCGCCGCTTTTCT 1549  
DB 2925 ACTCCCGTGGCCCGCCAGCCACATGAGACAGCCCTCGCCCTCGCCCGCCGCTTTTCT 2984  
QY 1550 CTGCTTTCTACCGACATGTGACCCCGACAGCCCTCGCCCGCCAGCTGCTCCCGGGC 1609  
DB 2985 CTGCTTTCTACCGACATGTGACCCCGACAGCCCTCGCCCGCCAGCTGCTCCCGGGC 3044  
QY 1610 AGTACTGGGACCTTCCCTGGGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1669  
DB 3045 AGTACTGGGACCTTCCCTGGGGGAGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3104  
QY 1670 GCCTGGGCGCTCAGTGGGACTGCTGCTCCACAGCTGCGGCTGACCTCAGAGGCGGAGG 1729  
DB 3105 GCCTGGGCGCTCAGTGGGACTGCTGCTCCACAGCTGCGGCTGACCTCAGAGGCGGAGG 3164  
QY 1730 CAGGAACCTGAGTGGGCGGCTGCTGCTGCTCAGGATGGTCTCGGGGCTCGCTGT 1789  
DB 3165 CAGGAACCTGAGTGGGCGGCTGCTGCTGCTCAGGATGGTCTCGGGGCTCGCTGT 3224  
QY 1790 CATCAAGACACCCCTTCCCGAGCTCACCACATCTCATCAGCAAGCGGAGGACTT 1849  
DB 3225 CATCAAGACACCCCTTCCCGAGCTCACCACATCTCATCAGCAAGCGGAGGACTT 3284  
QY 1850 GGCTCCCGCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAACCTCAACCTCC 1909  
DB 3285 GGCTCCCGCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGGAACCTCAACCTCC 3344  
QY 1910 CCCCTGCTCGGTTGGTGACAGAGGGGTGGGACAGGGGCGGGGCTTCCCGCTGTACAT 1969  
DB 3345 CCCCTGCTCGGTTGGTGACAGAGGGGTGGGACAGGGGCGGGGCTTCCCGCTGTACAT 3404  
QY 1970 ACCCTGCCATACCAACCCAGGATTAATTCCTGCTGCTGTTTATTTATTTATTTT 2029

DB 3405 ACCCTGCCATACCAACCCAGGATTAATTCCTGCTGCTGTTTATTTATTTT 3464  
QY 2030 TTGTTTTGATTTTTTAAAGAATTTTCATTTTAAAGCACAAAAA 2076  
DB 3465 TGTTTTGATTTTTTAAAGAATTTTCATTTTAAAGCAAAAAA 3511  
RESULT 7  
ABK84517  
ID ABK84517 standard; cDNA; 2907 BP.  
XX AC ABK84517;  
XX DT 14-AUG-2002 (first entry)  
XX Human cDNA differentially expressed in granulocytic cells #1088.  
DE Human; ss: granulocytic cell; DNA chip; bacterial infection;  
KW viral infection; parasitic infection; protozoal infection;  
KW fungal infection; sterile inflammatory disease; psoriasis;  
KW rheumatoid arthritis; glomerulonephritis; asthma; thrombosis;  
KW cardiac reperfusion injury; renal reperfusion injury; ARDS;  
KW adult respiratory distress syndrome; inflammatory bowel disease;  
KW Crohn's disease; ulcerative colitis; periodontal disease;  
KW granulocyte activation; chronic inflammation; allergy.  
XX Homo sapiens.  
XX WO200228999-A2.  
XX PD 11-APR-2002.  
XX PF 03-OCT-2001; 2001WO-US30821.  
XX PF -03-OCT-2000; 2000US-237189P.  
XX PR (GENE-) GENE LOGIC INC.  
XX PI Beazer-Barclay Y, Weissman SM, Yamaga S, Vockley J;  
XX WPI; 2002-435328/46.  
PT Detecting granulocyte activation by detecting differential expression  
PT of genes associated with granulocyte activation, which serves as  
PT diagnostic markers that is useful for monitoring disease states and  
PT drug toxicity  
XX Claim 1; SEQ ID No 1088; 114pp; English.  
PS The invention relates to detecting (M1) granulocyte (GC) activation  
CC (GCA), by detecting the level of expression of gene(s) (Gs) identified by  
CC DNA chip analysis as given in the specification, and comparing  
CC the expression level to an expression level in an unactivated  
CC GC, where differential expression of Gs is indicative of GCA.  
CC Also included are modulating (M2) GA by contacting GC with an agent  
CC that alters the expression of at least one gene in Gs; (2) screening (M3)  
CC for an agent capable of modulating GCA or an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease using the  
CC gene expression profile; (3) detecting (M4) an inflammation (especially  
CC chronic) in a tissue, an allergic response in a subject, exposure of a  
CC subject to a pathogen or sterile inflammatory disease, by detecting the  
CC level of expression in a sample of the tissue of gene(s) from Gs, where  
CC the level of expression of the gene is indicative of inflammation;  
CC (4) treating (M5) an inflammation (especially chronic) or in a tissue,  
CC an allergic response in a subject, exposure of a subject to a pathogen  
CC or sterile inflammatory disease, by contacting a tissue having  
CC inflammation with an agent that modulates the expression of gene(s)  
CC from Gs in the tissue. M1 is useful for detecting GCA; M2 is useful for  
CC modulating GA; M3 is useful for screening an agent capable of modulating  
CC GCA preferably in an inflammation in a tissue; M4 is useful for  
CC detecting an inflammation (especially chronic) in a tissue, an allergic

CC response in a subject, exposure of a subject to a pathogen or sterile  
CC inflammatory disease (e.g. psoriasis, rheumatoid arthritis, renal  
CC glomerulonephritis, asthma, thrombosis, cardiac reperfusion injury, renal  
CC reperfusion injury, ARDS, adult respiratory distress syndrome,  
CC inflammatory bowel disease, Crohn's disease, ulcerative colitis,  
CC peridontal disease; also bacterial infection, viral infection,  
CC parasitic infection, protozoal infection, fungal infection and M5 is  
CC useful for treating one of the above conditions. The present  
CC sequence represents a gene differentially expressed in granulocytes.  
CC Note: The sequence data for this patent did not form part  
CC of the printed specification, but was obtained in electronic  
CC format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ

Sequence 2907 BP; 561 A; 1011 C; 789 G; 546 T; 0 other;

Query Match 85.3%; Score 1778.8; DB 24; Length 2907;  
Best Local Similarity 99.3%; Pred. No. 0;  
Matches 1786; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
QY 288 CTCATTGAGACCCAGACAGCAGCTTCTGAAGAGATAGTCCCGAGCCCTCCCTGCCACCC 347  
DB 279 CACCATTTGAGACCCAGACAGCAGCTTCTGAAGAGATAGTCCCGAGCCCTCCCTGCCACCC 338  
QY 348 CCCTCTACCCCGCATCTACAAAGCTTCTTGTCTGTCTGAGGACAGTCCCTCAGGCTACCA 407  
DB 339 CCCTCTACCCCGCATCTACAAAGCTTCTTGTCTGTCTGAGGACAGTCCCTCAGGCTACCA 398  
QY 408 CTATGGGGTCAGCGCCCTGTGAGGCTCTCAAGGGCTTCTCCGCGCAGCATCCAGAGAA 467  
DB 399 CTATGGGGTCAGCGCCCTGTGAGGCTCTCAAGGGCTTCTCCGCGCAGCATCCAGAGAA 458  
QY 468 CATGGGTACAGTGTACCGGGACAGAGTGTGATCATCAACAGGTGACCCGAGACCC 527  
DB 459 CATGGGTACAGTGTACCGGGACAGAGTGTGATCATCAACAGGTGACCCGAGACCC 518  
QY 528 CTCGCGTACTGCGGACTGCAGAGTCTTGAAGTGGGATGTCCAGGAGTCTGTGAG 587  
DB 519 CTGCGAGTACTGCGGACTGCAGAGTCTTGAAGTGGGATGTCCAGGAGTCTGTGAG 578  
QY 588 AACGACCGAACAAGAGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 647  
DB 579 AACGACCGAACAAGAGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 638  
QY 648 GCTGACCGCGAGTGTGGGAGTCTATTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 707  
DB 639 GCTGACCGCGAGTGTGGGAGTCTATTGAGAGTGTGAGAGTGTGAGAGTGTGAGAGTGTGAG 698  
QY 708 CCCTGCGCTCTGCGAGTGTGGGCAATACACTACGAAACAGCTTCAGAAACAGCTGTCTC 767  
DB 699 CCCTGCGCTCTGCGAGTGTGGGCAATACACTACGAAACAGCTTCAGAAACAGCTGTCTC 758  
QY 768 TCTGGACATTGACCTCTGGGACAGTTCAGTGAATCTCCACCAAGTGTGATCATTAAGAC 827  
DB 759 TCTGGACATTGACCTCTGGGACAGTTCAGTGAATCTCCACCAAGTGTGATCATTAAGAC 818  
QY 828 TGTGGAGTTCGCAAGCAGTGTCCCGGGTTCACCAACCTTACCAATCGCGAGATACAC 887  
DB 819 TGTGGAGTTCGCAAGCAGTGTCCCGGGTTCACCAACCTTACCAATCGCGAGATACAC 878  
QY 888 CCTCTCAAGGCTGCGCTGTGACATCTGTATCTCGGATCTGACGCGGTACAGGCC 947  
DB 879 CCTCTCAAGGCTGCGCTGTGACATCTGTATCTCGGATCTGACGCGGTACAGGCC 938  
QY 948 CGAGCAGGACACCATGACCTTCTCGGACGGCTGACCTCAACCGGACCCAGATGCACAA 1007  
DB 939 CGAGCAGGACACCATGACCTTCTCGGACGGCTGACCTCAACCGGACCCAGATGCACAA 998  
QY 1008 CGCTGGCTTCGGCCCTCTACCGACCTGTGCTTTGCTTCGCCAACAGCTGTGCCCT 1067  
DB 999 CGCTGGCTTCGGCCCTCTACCGACCTGTGCTTTGCTTCGCCAACAGCTGTGCCCT 1058  
QY 1068 GGAGATGGATGATGCGGAGAGCGGGGTGCTCAGCGCCATCTGCGCTCATCTGCGGAGACCG 1127

DB 1059 GGAGATGGATGATGCGGAGAGCGGGCTGCTCAGCGCCATCTCCTCATCTGCGGAGACCG 1118  
QY 1128 CCAGGACCTGGAGCAGCGGACCGGGTGTGACATGTGAGGAGCGGCTGTCTGGAGCGCT 1187  
DB 1119 CCAGGACCTGGAGCAGCGGACCGGGTGTGACATGTGAGGAGCGGCTGTCTGGAGCGCT 1178  
QY 1188 AAAGTCTACGTGTGGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAT 1247  
DB 1179 AAAGTCTACGTGTGGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAT 1238  
QY 1248 GAAGTACTACTGCTGCGAAGCATCAGCCCAAGGGGTGAGCGGGTGTGATCAGCGTGA 1307  
DB 1239 GAAGTACTACTGCTGCGAAGCATCAGCCCAAGGGGTGAGCGGGTGTGATCAGCGTGA 1298  
QY 1308 GATGAGATCCCGGCTCCATGCGCGCTTCTCATCAGGAAATGTTGGAGAACTCAGAGGG 1367  
DB 1299 GATGAGATCCCGGCTCCATGCGCGCTTCTCATCAGGAAATGTTGGAGAACTCAGAGGG 1358  
QY 1368 CCTGGACACTCTGAGCGGACAGCGGGGTGGGGGGGAGCGGGGTGGGCTGCCCC 1427  
DB 1359 CCTGGACACTCTGAGCGGACAGCGGGGTGGGGGGGAGCGGGGTGGGCTGCCCC 1418  
QY 1428 CCCGCCAGGACAGTGTAGCCCGCAGCTCAGCCCGCAGCTCCAAACAGAGCAGCCCGGCCAC 1487  
DB 1419 CCCGCCAGGACAGTGTAGCCCGCAGCTCAGCCCGCAGCTCCAAACAGAGCAGCCCGGCCAC 1478  
QY 1488 CCACCTCCCGTACCGCCCGCAGCTGTGAGACAGCGCTTCCGCTCCGCGCCCGCTTTT 1547  
DB 1479 CCACCTCCCGTACCGCCCGCAGCTGTGAGACAGCGCTTCCGCTCCGCGCCCGCTTTT 1538  
QY 1548 CTCCTGCTTTTACCGGACATGTGACCCCGCAGCGCCCTGCCCGCCACCTGCCCTCCCG 1607  
DB 1539 CTCCTGCTTTTACCGGACATGTGACCCCGCAGCGCCCTGCCCGCCACCTGCCCTCCCG 1598  
QY 1608 GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1667  
DB 1599 GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658  
QY 1668 AGGCTTGGGCGCTCAGTGGAGTGCCTGCTCCACAGCTGGGCTGAGCTCAGAGCGGAG 1727  
DB 1659 AGGCTTGGGCGCTCAGTGGAGTGCCTGCTCCACAGCTGGGCTGAGCTCAGAGCGGAG 1718  
QY 1728 GCCAGGAATCAGTGGAGCGCCCTGCTGCTGCTGCTCAGATGGGCTCCTGGGGGCTCGTG 1787  
DB 1719 GCCAGGAATCAGTGGAGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1778  
QY 1788 TTCATCAAGACACCCCTCTGCGGAGTCCACCATCTTTCATCACCAGCAACCGCCAGGAC 1847  
DB 1779 TTCATCAAGACACCCCTCTGCGGAGTCCACCATCTTTCATCACCAGCAACCGCCAGGAC 1838  
QY 1848 TTGGCTCCCGCATCTCAGAACTCAGAGCGCTTCCCGCAGCTGGGGAACCTCAACCT 1907  
DB 1839 TTGGCTCCCGCATCTCAGAACTCAGAGCGCTTCCCGCAGCTGGGGAACCTCAACCT 1898  
QY 1908 CCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1967  
DB 1899 CCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1958  
QY 1968 ATACCTGCCATACCAACCGGAGTATTAATTTCTCGCTGGTGTGCTTTTATTTATTAATTT 2027  
DB 1959 ATACCTGCCATACCAACCGGAGTATTAATTTCTCGCTGGTGTGCTTTTATTTATTAATTT 2018  
QY 2028 TTTTGTGTTGATTTTTTAAAGAAATTTTCAATTTAAAGCAACAAAAA 2085  
DB 2019 TTTTGTGTTGATTTTTTAAAGAAATTTTCAATTTAAAGCAATTTTACTCAAGGAA 2076

RESULT 8  
ABL65214  
ID ABL65214 standard; DNA; 2907 BP.  
XX  
AC ABL65214;



XX	15-MAY-2002	(first entry)
DT		
XX		
XX		
XX		
KW	Lung cancer related gene sequence SEQ ID NO:3551.	
KW		
KW	Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;	
KW	stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;	
KW	cystosatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;	
KW	gene; ds..	
XX		
XX		
OS	Homo sapiens.	
XX		
PN	WO200194529-A2.	
XX		
XX	13-DEC-2001.	
XX		
XX		
XX	30-MAY-2001; 2001WO-US10838.	
XX		
PR	05-JUN-2000; 2000US-209473P.	
PR	05-JUN-2000; 2000US-209531P.	
PR	18-SEP-2000; 2000US-233133P.	
PR	18-SEP-2000; 2000US-233617P.	
PR	20-SEP-2000; 2000US-234009P.	
PR	20-SEP-2000; 2000US-234034P.	
PR	20-SEP-2000; 2000US-234052P.	
PR	22-SEP-2000; 2000US-234509P.	
PR	22-SEP-2000; 2000US-234567P.	
PR	25-SEP-2000; 2000US-234923P.	
PR	25-SEP-2000; 2000US-234924P.	
PR	25-SEP-2000; 2000US-235077P.	
PR	25-SEP-2000; 2000US-235082P.	
PR	25-SEP-2000; 2000US-235134P.	
PR	25-SEP-2000; 2000US-235280P.	
PR	26-SEP-2000; 2000US-235637P.	
PR	26-SEP-2000; 2000US-235638P.	
PR	27-SEP-2000; 2000US-235711P.	
PR	27-SEP-2000; 2000US-235720P.	
PR	27-SEP-2000; 2000US-235840P.	
PR	27-SEP-2000; 2000US-235863P.	
PR	28-SEP-2000; 2000US-236028P.	
PR	28-SEP-2000; 2000US-236032P.	
PR	28-SEP-2000; 2000US-236033P.	
PR	28-SEP-2000; 2000US-236034P.	
PR	28-SEP-2000; 2000US-236109P.	
PR	28-SEP-2000; 2000US-236111P.	
PR	29-SEP-2000; 2000US-236842P.	
PR	29-SEP-2000; 2000US-236891P.	
PR	02-OCT-2000; 2000US-237172P.	
PR	02-OCT-2000; 2000US-237173P.	
PR	02-OCT-2000; 2000US-237278P.	
PR	02-OCT-2000; 2000US-237294P.	
PR	02-OCT-2000; 2000US-237295P.	
PR	02-OCT-2000; 2000US-237316P.	
PR	03-OCT-2000; 2000US-237425P.	
PR	03-OCT-2000; 2000US-237598P.	
PR	03-OCT-2000; 2000US-237604P.	
PR	03-OCT-2000; 2000US-237606P.	
PR	03-OCT-2000; 2000US-237608P.	
PR	01-NOV-2000; 2000US-244867P.	
PR	01-NOV-2000; 2000US-245084P.	
XX		
XX	(AVAL-) AVALON PHARM.	
XX		
PI	Young PE, Augustus M, Carter KC, Ebner R, Endress G, Horrigan S;	
PI	Soppet DR, Weaver Z;	
XX		
XX	WPI: 2002-188264/24.	
XX		
PT	Screening for anti-neoplastic agent involves exposing cells to a	
PT	chemical agent to be tested for anti-neoplastic activity, and	
PT	determining a change in expression of a gene of a signature gene set	
XX		
PS	Claim 1; SEQ ID 3551; 44pp; English.	
PS		



```
QY 1008 CGCTGGCTTCCGCCCCCTCACCAGCTGCTGCTTGGCTTCCGCAACCAAGCTGCTGCCCT 1067
Db 999 CGCTGGCTTCCGCCCCCTCACCAGCTGCTGCTTGGCTTCCGCAACCAAGCTGCTGCCCT 1058
QY 1068 GGAGATGATGATCGGAGAGCGGGCTCTCAGGCGCCATCTGCCTCATCTCGGAGACCG 1127
Db 1059 GGAGATGATGATCGGAGAGCGGGCTCTCAGGCGCCATCTGCCTCATCTCGGAGACCG 1118
QY 1128 CCAGGACTTGAGAGCGCGGACCGGGTGACATCTGTCAGGAGCGCTGCTGGAGCGCT 1187
Db 1119 CCAGGACTTGAGAGCGCGGACCGGGTGACATCTGTCAGGAGCGCTGCTGGAGCGCT 1178
QY 1188 AAGGCTCTAGCTGAGGAGCGGCGCCAGCGCCCGCCACATGTTCCCAAGATCTAAT 1247
Db 1179 AAGGCTCTAGCTGAGGAGCGGCGCCAGCGCCCGCCACATGTTCCCAAGATCTAAT 1238
QY 1248 GAAGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGAGCGGTGATCAGCGTAA 1307
Db 1239 GAAGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGAGCGGTGATCAGCGTAA 1298
QY 1308 GATGGAGATCCCGGGCTCCATGCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGG 1367
Db 1299 GATGGAGATCCCGGGCTCCATGCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGG 1358
QY 1368 CCTGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGAGCGGGTGGCTGCCCTC 1427
Db 1359 CCTGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGAGCGGGTGGCTGCCCTC 1418
QY 1428 CCGCCAGGCGAGCTGTAGCCCGGAGCTCAGCGCCCGGAGCTCCACAGAGAGCGCGGCGAC 1487
Db 1419 CCGCCAGGCGAGCTGTAGCCCGGAGCTCAGCGCCCGGAGCTCCACAGAGAGCGCGGCGAC 1478
QY 1488 CCACTCCCGTACCAGCGCCACATGACACAGCGCTCGCTCGCTCCGCGCGCTTTT 1547
Db 1479 CCACTCCCGTACCAGCGCCACATGACACAGCGCTCGCTCGCTCCGCGCGCTTTT 1538
QY 1548 CTCTGCTTTTACCAGCACTGACCGCCCGGAGCGCGCTCGCTCGCTCGCTCGCTCG 1607
Db 1539 CTCTGCTTTTACCAGCACTGACCGCCCGGAGCGCGCTCGCTCGCTCGCTCGCTCG 1598
QY 1608 GCAGTACTGGGAGCTTCCCTGGGGGAGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1667
Db 1599 GCAGTACTGGGAGCTTCCCTGGGGGAGCGGGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658
QY 1668 AGGCTTGGGCGCTCAGTGGACTGCTTCCCGGAGCGCGCTCGCTCGCTCGCTCGCTCG 1727
Db 1659 AGGCTTGGGCGCTCAGTGGACTGCTTCCCGGAGCGCGCTCGCTCGCTCGCTCGCTCG 1718
QY 1728 GCGAGAACTGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1787
Db 1719 GCGAGAACTGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1778
QY 1788 TTCATCAGACACCTCTGCGCCAGCTCACCAGCTTCCATCACCAGCAACCCAGGAC 1847
Db 1779 TTCATCAGACACCTCTGCGCCAGCTCACCAGCTTCCATCACCAGCAACCCAGGAC 1838
QY 1848 TTGGCTTCCCGCTCTCAGAACTCAGAGCCATGCTCCCGAGCTGGGGAACCTCAACCT 1907
Db 1839 TTGGCTTCCCGCTCTCAGAACTCAGAGCCATGCTCCCGAGCTGGGGAACCTCAACCT 1898
QY 1908 CCCCCCTGCCCTGGTGTGACAGAGGGGGTGGGAGCGGGGGGGGGGGGGGGGGGGGGGGGG 1967
Db 1899 CCCCCCTGCCCTGGTGTGACAGAGGGGGTGGGAGCGGGGGGGGGGGGGGGGGGGGGGGGG 1958
QY 1968 ATACCTGCCATACCAACCCAGGTATTAAATCTCGCTGGTTTGTATTTTAAATTT 2027
Db 1959 ATACCTGCCATACCAACCCAGGTATTAAATCTCGCTGGTTTGTATTTTAAATTT 2018
QY 2028 TTTTGTGTTTGTATTTTAAATTAAGAAATTTTCAATTTTAAAGCAAAAAA 2085
Db 2019 TTTTGTGTTTGTATTTTAAATTAAGAAATTTTCAATTTTAAAGCAATTTTAAAGCA 2076
```

```
RESULT 9
AAV64991
ID AAV64991 standard; cDNA; 2928 BP.
XX
AC AAV64991;
XX
DT 05-FEB-1999 (first entry)
XX
DE Human RAR-alpha cDNA.
XX
KW Fusion protein; myl; retinoic acid receptor-alpha; RAR; human;
KW acute promyelocytic leukaemia; APL; t(15;17); translocation;
KW treatment; all-trans retinoic acid; ss.
XX
OS Homo sapiens.
XX
Key Location/Qualifiers
CDS 103..1491
FT FT
FT /*tag= a
FT /product= "RAR-alpha"
FT /note= "retinoic acid receptor-alpha"
XX
PN US5843642-A.
XX
PD 01-DEC-1998.
XX
PF 21-JUL-1993; 93US-0095728.
XX
PR 21-JUL-1993; 93US-0095728.
PR 22-MAR-1991; 91US-0673838.
PR 22-MAR-1991; 91US-0675084.
XX
XX (SLOK ) SLOAN KETTERING INST CANCER RES.
XX
PI Dmitrovsky E, Frankel S, Miller WH, Warrell RP;
XX
WPI: 1999-044563/04.
P-PSDB; AAW81984.
XX
Diagnosis of acute promyelocytic leukaemia - by detecting nucleic
acid encoding abnormal retinoic acid receptor-alpha
XX
PS Disclosure; Column 29-34; 38pp; English.
XX
CC This sequence encodes the human retinoic acid receptor alpha, RAR-alpha
CC which is used in a method for identifying a subject with acute
CC promyelocytic leukaemia (APL) resulting from a t(15;17) translocation who
CC will respond to treatment with all-trans retinoic acid. The protein can
CC also be used to identify a subject with indications of APL who will not
CC respond to treatment with all-trans retinoic acid.
XX
SQ Sequence 2928 BP; 582 A; 1011 C; 789 G; 546 T; 0 other;
```

```
Query Match 85.3%; Score 1778.8; DB 20; Length 2928;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1786; Conservative 0; Mismatches 12; Indels 0; Gaps 0;
QY 288 CTCATTGAGACCCAGAGCAGCTTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACC 347
Db 279 CACCATTGAGACCCAGAGCAGCTTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACC 338
QY 348 CCTCTACCCCGCATCTACAAGCCTTGTCTGTGTCAGGACAAGTCTCAGGCTACCA 407
Db 339 CCTCTACCCCGCATCTACAAGCCTTGTCTGTGTCAGGACAAGTCTCAGGCTACCA 398
QY 408 CTATGGGCTCAGGCGCTGTGAGGGCTGCAAGGCTTCTTCCCGCGCAGCATCCAGAGAA 467
Db 399 CTATGGGCTCAGGCGCTGTGAGGGCTGCAAGGCTTCTTCCCGCGCAGCATCCAGAGAA 458
QY 468 CATGTTGTACAGTGTCTACCGGAGCAAGACTGCATCATCAACAGGTGACCCGGAACCC 527
Db 459 CATGTTGTACAGTGTCTACCGGAGCAAGACTGCATCATCAACAGGTGACCCGGAACCC 518
```



XX (INNO-) INNOVIR LAB INC.  
PA George ST, Goldberg AR, Pace U;  
PI WPI: 1996-300650/30.  
XX RNA construct(s) including ribozyme(s) and antisense  
XX oligo:nucleotide(s) - for the inactivation of RNA associated with,  
XX e.g. promyelocytic leukemia or follicular lymphoma  
PS Example 1; Page 53-55; 81pp: English.  
XX PML-RAR-alpha RNA (AAT33246) is associated with acute promyelocytic  
CC leukaemia (APL). APL is characterised by a balanced and  
CC reciprocal translocation between the long arms of chromosomes 15  
CC and 17, resulting in a fusion of the retinoic acid receptor gene  
CC (RAR-alpha, see also AAT33260) and a gene for a putative  
CC transcription factor, PML. Methods for treating APL involve the  
CC use of ribozymes, external guide sequences and antisense  
CC oligonucleotides (see also AAT33245 and AAT33248-57) that specifically  
CC cleave the PML-RAR-alpha fusion mRNA (see also AAT33251) but not  
CC wild-type RAR-alpha mRNA.  
XX Sequence 3511 BP; 713 A; 1168 C; 1038 G; 591 U; 1 other;  
SQ

Query Match 85.2%; Score 1777.4; DB 17; Length 3511;  
Best Local Similarity 81.5%; Pred. No. 0;  
Matches 1456; Conservative 325; Mismatches 6; Indels 0; Gaps 0;

QY 290 CCATTGAGACCCAGACGACGCTTCGAGAGATAGTCCCGAGCCCTCCCTCCGACCCC 349  
DB 1725 CCAUUGAGACCCAGACGACGCUUUGAAGAGAGAGUGGCCAGCCUCCUCCGACCCC 1784  
QY 350 CTCCTACCCGCGCTACAAAGCCTTGCTGTCTGTCAGGACAGTCTCTAGGCTACCACT 409  
DB 1785 CUCUACCCGCAUCACAAAGCCUUGUUGUGUCAGGACGACGACGACGACGACGAC 1844  
QY 410 ATGGGGTACGCGCTGTAGGGCTGCAAGGCTTCTTCGCGCGCAGCATCCAGAAACA 469  
DB 1845 AUGGGGUCACGCGGUGAGGCGUGCAAGGCGUUCUUCGCGCGCAGCAUCCAGAAACA 1904  
QY 470 TGGGTACAGCTGTCACCGGACAACTGTCATCATCAACAGGTGACCGGACCCCT 529  
DB 1905 UGGUGACACGUGUACCCGGGACAAAGCAUACAUCAACAGAGGACCGGACCGGAC 1964  
QY 530 GCAGTACTCCGAGTGCAGAAAGTGTCTTGAAGTGGGCTGTCCAAAGGAGTCTGTGAGAA 589  
DB 1965 GCCAGUACGCGACUGCAGAGUGCUUUGAAGUGGCGCAUGUCCAGAGGACUGAGAA 2024  
QY 590 ACGACCGAAACAAAGAAAGAGAGGAGTGCACAAAGCCCGAGTGTCTGTAGAGTACACGC 649  
DB 2025 ACGACCGAAACAAAGAAAGAGAGGAGGUGGCCAAAGCCCGAGGACGACGACGAC 2084  
QY 650 TCACCCGAGGTGGGGAGCTCATGAGAGGTGCGCAAGGCGCAAGCGACGAAACCTTC 709  
DB 2085 UGACCCGAGGAGGUGGAGGAGCUUUGAAGGUGGCGCAAGGCGACCGAGAAACCUUC 2144  
QY 710 CTGCCCTCTGCCAGCTGGGCAATACACTACGAAACACAGCTCAGAAACACGCTGTCTC 769  
DB 2145 CUGCCUUCGCGAGCGUGGCAAAUACACUACGAAACAGCUCAGAAACAGGUGUCUC 2204  
QY 770 TGGACATTCACCTCTGGGACAAAGTTGAGTGAAGTCTCCACCAAGTGCATCAATTAAGT 829  
DB 2205 UGGACAUUAGACCCUUGGACAAAGUUCAGUGAAGUUCUCCACCAAGUGCAUCAUUAAGAC 2264  
QY 830 TGGAGTTCGCCAGGAGCTGCCGCTTACCAACCTGACCATCGCGGACCATCACCC 889  
DB 2265 UGGAGUUGCCCAAGGACGUGCGGCUUCCACCAACCCUCCACCAUCCGCGGACGACGAC 2324  
QY 890 TCCTCAAGGCTCCCTGCTGGACATCCTGATCTGCGGATCTGACGCGGTACACGCCCG 949  
DB 2325 UCCUCAAGGCGCGUCCUGGACAUCCUUGUUGGAGUUGGCGGAGUUGCAGCGGACGCGCG 2384

QY 950 AGCAGGACACCATGACCTTCTCGACGGCTGACCTGAACCGGACCCAGATGCAACAG 1009  
DB 2385 AGCAGGACACCAUGACCUUCGCGGGUGAGCCUAGACCGGACCCAGACCAACAG 2444  
QY 1010 CTGGCTTCGGCCCCCTCACCGACCTGCTTTCCTTCGCGCAACAGCTGTGCCCCCTGG 1069  
DB 2445 CUGGCUUGCGCCCCUCACCGACCGUGUUCUUGCGCCCAACAGCUGCUGCCCCUGG 2504  
QY 1070 AGATGATGATGCGGAGACGGGCTGTACGCGCATCTGCTCATCTGCGGAGACCGCC 1129  
DB 2505 AGAUGGAUGAUGCGGAGACGGGCGUCAGCGCAUCUGCCUACUUGCGGAGACCGCC 2564  
QY 1130 AGGACCTGGAGACGACCGGCGGTGACATGCTGCAGGAGCCCTGCTGAGGCGCTAA 1189  
DB 2565 AGGACCUUGGAGACGACCGGCGGTGACATGCTGCAGGAGCCCGCUGGAGGCGCUAA 2624  
QY 1190 AGTCTACGTGCGGAAGCGGAGCGCCAGCGCCCGCCACATCTTCCCAAGATGCTAATGA 1249  
DB 2625 AGGUCUACGUGGGAAGCGGAGCGCCAGCGCCCGCCACAUUGUCCCAAGAUUGCUAUGA 2684  
QY 1250 AGATTACTGACCTGCGAAGCATCAGCCCGCAAGGGGGTGTAGCGGTGATCATCGTGAAGA 1309  
DB 2685 AGAUUACUGACCUUGCGAAGCAUCAAGCGCCAAAGGGGCGUGAGCGGUGAUCAGCUAAGA 2744  
QY 1310 TGGAGATCCCGGGCTCCATGCGCGCTCTCATCCAGAAATTTGGAGAACTCAGAGGCC 1369  
DB 2745 UGGAGAUCCCGGGCGUCCAGUCCGCGCUCUCCAUCCAGGAAAUUGGAGAAACUCAGAGGCC 2804  
QY 1370 TGGACACTGTAGCGGACGACCGGGGGGTGGGGGCGGAGCGGGGTGCGCTGCCCGCCC 1429  
DB 2805 UGGACACUCUGAGCGGACGACCGGGGGGUGGGGGGCGGAGCGGGGUGGCGUCCGCCCCC 2864  
QY 1430 CGCAGGACGCTGTAGCCCCAGCCTCAGCCCCAGCTTCAACAGAAAGACGCGCGCCACCC 1489  
DB 2865 CGCAGGACGUGUAGUCCCGAGCCAGCCAGCCAGCUCUCCAAACAGAGACGCGCGCCACCC 2924  
QY 1490 ACTCCCGGTACCGCCCGACGCGCATGGACACAGCCCTCGCCCTCGCGCCCGCTTTCT 1549  
DB 2925 ACUCCCGUGACCGCCCGACCGCAUUGGACACAGCCCGCGCGCGCGCGUUCU 2984  
QY 1550 CTCGCTTTCACGACCATGTGACCCCGACAGCCCTGCCCCACCTGCCCCCTCCCGGGC 1609  
DB 2985 CUGCCUUCUACCGACCAUGGAGCCCGACCGCCAGCCUCCUCCCGCGCGCGCGCGCG 3044  
QY 1610 AGTACTGGGACCTTCCCTGGGGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1669  
DB 3045 AGUACUGGGACCUUCCUUGGGGACGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3104  
QY 1670 GCTTGGGCGCTCAGTGGACTGCTTCCACAGCCCTGGGCTGGGCTCAGTCCAGAGGCGGAG 1729  
DB 3105 GCGUGGCGCGUACUGGACUGCCUCCACAGCCGCGGCGGCGGCGGCGGCGGCGGAGG 3164  
QY 1730 CAGGAAGTGTAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1789  
DB 3165 CAGGAACUGAGUGAGGCG 3224  
QY 1790 CATCAAGACACCCCTCTGCGCGAGCTCACCATCTTCTATCATCAGCAACCGCAGGACTT 1849  
DB 3225 CAUCAAGACACCCUCCUCCCGACGACCAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 3284  
QY 1850 GGCTCCCGCATCTCAGAACTCAGAACCTTGTCTCCCGAGCTGGGGAACCTCAACCTCC 1909  
DB 3285 GCGUCCCGCAUCCUAGAACUACAAGCCAUUGUCCCGAGCUGGGGGAACCCUACACCC 3344  
QY 1910 CCCTGCTCGTGTGTGTGACAGAGGGGTGGACAGGCGGGGGTTCGCCCTGTACAT 1969  
DB 3345 CCCUGCGCGGUGUGACAGAGGGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 3404  
QY 1970 ACCTGCTCATCAACCGCCAGGTAATCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2029  
DB 3405 ACCUGGCAUACCAACCGCCAGGUAUUAUUAUUAUUAUUAUUAUUAUUAUUAUUAU 3464

QY 2030 TTGTTTGTATTTTAAATAAGATTTTCATTTTAAGCACAATAAAA 2076  
? : :::::|||||:|||||:|||||:|||||:|||||:|||||:|||||  
Db 3465 UGUUUUGAUUUUUUAUAGAUAUUUUCAUUUAUUAUUAUUAUUA 3511

## RESULT 11

AAQ29338  
ID AAQ29338 standard; cDNA; 2928 BP.

AC AAQ29338;

DT 09-MAR-1993 (first entry)

DE RAR-alpha gene.

KW Retinoic acid receptor; RAR-alpha; myl; acute promyelocytic leukemia;  
APL; translocation; chromosome 17; chromosome 15; PCR; primer;  
[t(15;17)(q21;q11-22)]; breakpoint; polymerase chain reaction; ss.

OS Synthetic.

XX Location/Qualifiers  
FH Key 103..1488  
FT CDS /\*tag= a

XX WO9216660-A.

PD 01-OCT-1992.

PF 23-MAR-1992; 92WO-US02320.

PR 22-MAR-1991; 91US-0673838.

PR 22-MAR-1991; 91US-0675084.

XX (SLOK ) SLOAN KETTERING INST CANCER.

XX Dmltrovsky E, Evans RM, Frankel S, Kazizuka A, Miller WH;  
PI Warrell RP;

DR WPI; 1992-349240/42.

DR P-PSDB; AAR27534.

XX Marker for acute promyelocytic leukaemia and other neoplasias -  
PT comprising nucleic acid and encoded abnormal retinoic acid  
PT receptor-alpha receptor

XX Disclosure; Page 43-46; 84pp; English.

XX The sequence given represents the nucleic acid sequence of the  
CC retinoic acid receptor (RAR)-alpha gene. This gene is disrupted in  
CC a translocation of a portion of the long arm of chromosome 17 onto  
CC the long arm of chromosome 15 [t(15;17)(q21;q11-22)]. This causes a  
CC fusion between RAR-alpha and myl which is characteristic of acute  
CC promyelocytic leukemia (APL). The breakpoint region has been cloned  
CC and it has been shown that DNA rearrangements are clustered in the  
CC region of the first intron of RAR-alpha. This sequence was isolated  
CC by polymerase chain reaction (PCR). The primers used for amplification  
CC of this sequence can also be used to amplify the translocated region.

XX Sequence 2928 BP; 582 A; 1012 C; 789 G; 545 T; 0 other;

Query Match 85.2%; Score 1777.2; DB 13; Length 2928;  
Best Local Similarity 99.3%; Pred. NO. 0;  
Matches 1785; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 288 CTCATTGAGACCCAGAGCAGCAGTTCTGAAGAGATAGTGCCAGCCCTCCCTCGCCACC 347

Db 279 CACCATTGAGACCCAGAGCAGCAGTTCTGAAGAGATAGTGCCAGCCCTCCCTCGCCACC 338

QY 348 CCCTTACCCCGCATCTACAAGCCTTGCTTGTCTGTCAGGACAAGTCTCAGGCTACCA 407

Db 339 CCCTTACCCCGCATCTACAAGCCTTGCTTGTCTGTCAGGACAAGTCTCAGGCTACCA 398

QY 408 CTATGGGTTCAGCCCTGTGAGGGCTCAAGGGCTTCTTCGCGCGCAGCATCCAGAAGAA 467  
Db 399 CTATGGGTTCAGCCCTGTGAGGGCTCAAGGGCTTCTTCGCGCGCAGCATCCAGAAGAA 458  
QY 468 CATGGTGTACACGTGTACCCGGGACAAAGAACTGCATCATCAACAAGGTGACCCGGAACCC 527  
Db 459 CATGGTGTACACGTGTACCCGGGACAAAGAACTGCATCATCAACAAGGTGACCCGGAACCC 518  
QY 528 CTGCCAGTACTCGCGAGTGCAGAGTCTTTGAAGTGGGATGTCCTCAAGGAGTCTGTGAG 587  
Db 519 CTGCCAGTACTCGCGAGTGCAGAGTCTTTGAAGTGGGATGTCCTCAAGGAGTCTGTGAG 578  
QY 588 AAACGACCCGAACAAGAAAGAGAGGTGCCCAAGCCCGAGTCTCTGAGAGCTACAC 647  
Db 579 AAACGACCCGAACAAGAAAGAGAGGTGCCCAAGCCCGAGTCTCTGAGAGCTACAC 638  
QY 648 GCTGACGCCGAGGTGGGAGGTCTATGAGAGGTGGCAAGCGCACCGAGGAACCTT 707  
Db 639 GCTGACGCCGAGGTGGGAGGTCTATGAGAGGTGGCAAGCGCACCGAGGAACCTT 698  
QY 708 CCCTGCCCTCTGCCAGCTGGGCAAAATACACTAGCAACAAGCTCAGAACACGTGTCTC 767  
Db 699 CCCTGCCCTCTGCCAGCTGGGCAAAATACACTAGCAACAAGCTCAGAACACGTGTCTC 758  
QY 768 TCTGGACATTGACCTCTGGGACAAAGTTTCACTGAACTCTCCACCAAGTGCATTTAAGAC 827  
Db 759 TCTGGACATTGACCTCTGGGACAAAGTTTCACTGAACTCTCCACCAAGTGCATTTAAGAC 818  
QY 828 TGTGGAGTTTCCCAAGCAGCTGCCCGCTTACACACCTTCCACCATCGCCGACCATCAC 887  
Db 819 TGTGGAGTTTCCCAAGCAGCTGCCCGCTTACACACCTTCCACCATCGCCGACCATCAC 878  
QY 888 CTTCTCAAGGTGCTGCTGGACATCTGATCTCTGGGATCTGACGCGGTACACGCC 947  
Db 879 CTTCTCAAGGTGCTGCTGGACATCTGATCTCTGGGATCTGACGCGGTACACGCC 938  
QY 948 CGAGCAGGACACCATGACCTTCTCGGACGGCTGACCTGAAACGGACCCAGATGCACAA 1007  
Db 939 CGAGCAGGACACCATGACCTTCTCGGACGGCTGACCTGAAACGGACCCAGATGCACAA 998  
QY 1008 CGCTGGCTTCGGCCCTTACCAGCTTGTTCCTTCGCTTCCCAACCATGCTGCTGCCCT 1067  
Db 999 CGCTGGCTTCGGCCCTTACCAGCTTGTTCCTTCGCTTCCCAACCATGCTGCTGCCCT 1058  
QY 1068 GGAGATGGATGTCGGAGACGGGGTGTCTAGCGGCATCTGCTCATCTCGGAGACCG 1127  
Db 1059 GGAGATGGATGTCGGAGACGGGGTGTCTAGCGGCATCTGCTCATCTCGGAGACCG 1118  
QY 1128 CCAGGACCTGGAGACCGCGGCTGACATGCTGCAGGAGCGCTGCTGGAGGCGCT 1187  
Db 1119 CCAGGACCTGGAGACCGCGGCTGACATGCTGCAGGAGCGCTGCTGGAGGCGCT 1178  
QY 1188 AAAGGTCTAGCTCGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAT 1247  
Db 1179 AAAGGTCTAGCTCGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAT 1238  
QY 1248 GAAGATTACTGACCTCGGAAGCATACGCGCAAGGGGGTGTAGCGGTGTATCAGCTGAA 1307  
Db 1239 GAAGATTACTGACCTCGGAAGCATACGCGCAAGGGGGTGTAGCGGTGTATCAGCTGAA 1298  
QY 1308 GATGGAGATCCCGGGTCCATGCGGCTCTCATCCAGGAATGTTGGAGAATCAGAGG 1367  
Db 1299 GATGGAGATCCCGGGTCCATGCGGCTCTCATCCAGGAATGTTGGAGAATCAGAGG 1358  
QY 1368 CTTGGACACTCTCAGCGGACAGCCGGGGTGGGGGCGGAGCGGGTGGCTGCCGCC 1427  
Db 1359 CTTGGACACTCTCAGCGGACAGCCGGGGTGGGGGCGGAGCGGGTGGCTGCCGCC 1418  
QY 1428 CCGCGCAGGCGAGCTGTAGCGCCAGCTCAGCGCCAGCTTCCAAACAGACGCCGGCCAC 1487  
Db 1419 CCGCGCAGGCGAGCTGTAGCGCCAGCTCAGCGCCAGCTTCCAAACAGACGCCGGCCAC 1478  
QY 1488 CCACCTCCCGGTGACCGCCACAGCCACATGAGACAGACCCCTCGGCCCTCGGCCCGGCTTTT 1547



```
Db 1059 GGAGTGGATGATCGGAGACGGGGCTGCTACGGCCATCTGCTCATCTCGCGAGACCG 1118
Qy 1128 CCAGGACCTGGAGACCGCGGAGCGGTGACATGCTGAGGAGCCGCTGCTGAGAGCGCT 1187
Db 1119 CCAGGACCTGGAGACCGCGGAGCGGTGACATGCTGAGGAGCCGCTGCTGAGAGCGCT 1178
Qy 1188 AAAGGTCTACGTGCGGAAGCGGAGGCCAGCGGCCCGCCACATGTTCCCAAGATGCTAAT 1247
Db 1179 AAAGGTCTACGTGCGGAAGCGGAGGCCAGCGGCCCGCCACATGTTCCCAAGATGCTAAT 1238
Qy 1248 GAAGATTACTGACCTCGGAAGCATCAGCGCCCAAGGGGGCTGAGCGGGTGTATCACGCTGAA 1307
Db 1239 GAAGATTACTGACCTCGGAAGCATCAGCGCCCAAGGGGGCTGAGCGGGTGTATCACGCTGAA 1298
Qy 1308 GATGAGATCCCGGGTCCATCGCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGG 1367
Db 1299 GATGAGATCCCGGGTCCATCGCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGG 1358
Qy 1368 CTTGACACTCTGAGCGGACAGCGGGGGGTGGGGGGCGGAGCGGGGTGGCCCTGCCGCC 1427
Db 1359 CTTGACACTCTGAGCGGACAGCGGGGGGTGGGGGGCGGAGCGGGGTGGCCCTGCCGCC 1418
Qy 1428 CCGCCAGGACGCTGTAGCCCCAGCTCAGCCCCAGCTCCAACAGAGAGCAGCCCGCCAC 1487
Db 1419 CCGCCAGGACGCTGTAGCCCCAGCTCAGCCCCAGCTCCAACAGAGAGCAGCCCGCCAC 1478
Qy 1488 CCACCTCCCGGTGACCGCCCGCCAGCCATGACACAGCCCTCGCCCTCCCGCGGCTTTT 1547
Db 1479 CCACCTCCCGGTGACCGCCCGCCAGCCATGACACAGCCCTCGCCCTCCCGCGGCTTTT 1538
Qy 1548 CTCTGCTTTTACCGACCATGTGACCCCGCCAGCCCTGCCCCACCTGCCCTCCCGG 1607
Db 1539 CTCTGCTTTTACCGACCATGTGACCCCGCCAGCCCTGCCCTCCCGG 1598
Qy 1608 GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1667
Db 1599 GCAGTACTGGGACCTTCCCTGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1658
Qy 1668 AGGCTTGGGCTTCACTGAGTGGCTGCTGCTCCACAGCCCTGGGCTGACGTCAGAGGCGGAG 1727
Db 1659 AGGCTTGGGCTTCACTGAGTGGCTGCTGCTCCACAGCCCTGGGCTGACGTCAGAGGCGGAG 1718
Qy 1728 GCCAGAACTAGTGGAGCCCTTGCTGGGTCTGAGGATGGGTCTCGGGGCTCGGTG 1787
Db 1719 GCCAGAACTAGTGGAGCCCTTGCTGGGTCTGAGGATGGGTCTCGGGGCTCGGTG 1778
Qy 1788 TTCATCAAGACACCCCTCTGCGGAGCTCAGCTACACATCTTATCACCAGCAAGCCAGGAC 1847
Db 1779 TTCATCAAGACACCCCTCTGCGGAGCTCAGCTACACATCTTATCACCAGCAAGCCAGGAC 1838
Qy 1848 TTGGCTCCCGCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGAACTCAACCT 1907
Db 1839 TTGGCTCCCGCATCTCAGAACTCACAAGCCATTGCTCCCGAGCTGGGAACTCAACCT 1898
Qy 1908 CCGGCTGCTCGGTGTGTGACAGAGGGGTGGGAGAGGGGGGGGGTTCGCCCTGTAC 1967
Db 1899 CCGGCTGCTCGGTGTGTGACAGAGGGGTGGGAGAGGGGGGGGGTTCGCCCTGTAC 1958
Qy 1968 ATACCTGCCATACCAACCCAGGTATTAATTTCTCGCTGTTTGTATTTAATTT 2027
Db 1959 ATACCTGCCATACCAACCCAGGTATTAATTTCTCGCTGTTTGTATTTAATTT 2018
Qy 2028 TTTTGTGTTGATTTTATTAAGATTTTCAATTTTAAGCACAAAAAATAAAAAA 2085
Db 2019 TTTTGTGTTGATTTTATTAAGATTTTCAATTTTAAGCACATTTATCTAAGGAA 2076
```

RESULT 13

AAQ81476

ID AAQ81476 standard; cDNA to mRNA; 2940 BP.

XX

AC AAQ81476;

```
XX 01-SEP-1995 (first entry)
DT
XX RAR-alpha clone lambda-KIR.
DE
XX RAR-alpha; retinoic acid receptor alpha; hematopoietic; stem cell;
KW differentiation; ds.
KW
XX Homo sapiens.
OS
XX
FH Key Location/Qualifiers
CDS 103..1491
FT /*tag= a
XX
PN W09504143-A.
XX
PD 09-FEB-1995.
XX
PF 28-JUL-1994; 94WO-US08450.
XX
PR 28-JUL-1993; 93US-0099242.
XX (HUTC-) HUTCHINSON CANCER RES CENT FRED.
XX
PI Collins SJ, Tsai S;
XX
DR P-PSDB; AAR68023.
XX
PT New haematopoietic stem cell lines with specific differentiation
PT properties - made by transfected stem cells with nucleic acid
PT encoding dominant negative suppressor of the retinoic acid
PT receptor alpha, useful e.g. for haematopoietic reconstitution
XX
PS Disclosure; Page 55-56; 100pp; English.
XX
CC An RAR-alpha cDNA (given in AAQ81477) was prepared that contained a
CC truncation of sequences coding for the C-terminal 59 amino acids
CC and part the 3' UTR of the wild-type human sequence (AAQ81476).
CC This truncated cDNA, designated RAR-alpha-403, encodes a protein
CC (AAR68024) containing the N-terminus, DNA-binding domain and part of
CC the hormone-binding domain of RAR-alpha (AAR68023).
XX
SQ Sequence 2940 BP; 594 A; 1012 C; 788 G; 546 T; 0 other;
Query Match 85.2%; Score 1777.2; DB 16; Length 2940;
Best Local Similarity 99.3%; Pred. No. 0;
Matches 1785; Conservative 0; Mismatches 13; Indels 0; Gaps 0;
Qy 288 CTCATTGAGACCCAGAGCAGCAGTTCCTGAAGAGATAGTGCACGCCCTCCCTCGCCACC 347
Db 279 CACCATGAGACCCAGAGCAGCAGTTCCTGAAGAGATAGTGCACGCCCTCCCTCGCCACC 338
Qy 348 CCCTCTACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGACAAAGTCTCTAGGCTACCA 407
Db* 339 CCCTCTACCCCGCATCTACAAGCCTTGTCTGTCTGTCAGACAAAGTCTCTAGGCTACCA 398
Qy 408 CTATGGGTGACGCCCTGTGAGGGCTGCAAGGGCTTCTTCGGCGGACATCCAGAAGAA 467
Db 399 CTATGGGTGACGCCCTGTGAGGGCTGCAAGGGCTTCTTCGGCGGACATCCAGAAGAA 458
Qy 468 CATGGGTGACACCTGTGACCGGACAAAGAACTCATCATCAACAAGGTGACCCGGAACCC 527
Db 459 CATGGGTGACACCTGTGACCGGACAAAGAACTCATCATCAACAAGGTGACCCGGAACCC 518
Qy 528 CTGCCAGTACTGCGGACTGCGAGAAGTGTCTTGAAGTGGGATGTCTCAAGGAGTCTGTGAG 587
Db 519 CTGCCAGTACTGCGGACTGCGAGAAGTGTCTTGAAGTGGGATGTCTCAAGGAGTCTGTGAG 578
Qy 588 AAACGACCGAAACAAGAAGAAGAGAGGTGCCCAAGCCCGAGTCTCTGAGAGCTACAC 647
Db 579 AAACGACCGAAACAAGAAGAAGAGAGGTGCCCAAGCCCGAGTCTCTGAGAGCTACAC 638
```



QY 648 GCTGACGCCGAGGTGGGGAGCTCATTGAGAAGTGGCGAAAGCGCACAGGAAACCTT 707  
Db 639 GCTGACGCCGAGGTGGGGAGCTCATTGAGAAGTGGCGAAAGCGCACAGGAAACCTT 698  
QY 708 CCCTGCCCCCTGCGCAGCTGGGCAAAATACACTACGAACAAACAGCTCAGAACACCTGTCTC 767  
Db 699 CCCTGCCCCCTGCGCAGCTGGGCAAAATACACTACGAACAAACAGCTCAGAACACCTGTCTC 758  
QY 768 TCTGGACATTGACCTCTGGGCAAGTTCAGTGAAGTCTCCACCAAGTGCATTAAGAC 827  
Db 759 TCTGGACATTGACCTCTGGGCAAGTTCAGTGAAGTCTCCACCAAGTGCATTAAGAC 818  
QY 828 TGTGGAGCTTGGCCCAAGCAGCTGCCCCGCTTCAACACCTCACCATTGCGCGACGATGAC 887  
Db 819 TGTGGAGCTTGGCCCAAGCAGCTGCCCCGCTTCAACACCTCACCATTGCGCGACGATGAC 878  
QY 888 CTCTCTCAAGGCTGCCCTCGCTGGACATCTCTGATCCTGGGATCTGCAAGCGGTACACGCC 947  
Db 879 CTCTCTCAAGGCTGCCCTCGCTGGACATCTCTGATCCTGGGATCTGCAAGCGGTACACGCC 938  
QY 948 CGAGCAGGACACCATGACCTTCTCGGAGCGGCTGACCTGACCGGACCGACCGAGTGCACAA 1007  
Db 939 CCAGCAGGACACCATGACCTTCTCGGAGCGGCTGACCTGACCGGACCGACCGAGTGCACAA 998  
QY 1008 CGCTGGCTTCGGCCCCCTCAACCGACCTGCTTTGCTTCCGCAACAGCTGCTGCCCT 1067  
Db 999 CGCTGGCTTCGGCCCCCTCAACCGACCTGCTTTGCTTCCGCAACAGCTGCTGCCCT 1058  
QY 1068 GGAGATGGATGATCGGAGAGCGGGCTGCTCAGCGCCATCTGCTCATCTCGCGAGACCG 1127  
Db 1059 GGAGATGGATGATCGGAGAGCGGGCTGCTCAGCGCCATCTGCTCATCTCGCGAGACCG 1118  
QY 1128 CCAGGACCTTGAGCAGCGCGGAGCGGCTGACATCTGCGAGAGCGGCTGCTGGAGCGCT 1187  
Db 1119 CCAGGACCTTGAGCAGCGCGGAGCGGCTGACATCTGCGAGAGCGGCTGCTGGAGCGCT 1178  
QY 1188 AAAGGTCTAGCTGCGGAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1247  
Db 1179 AAAGGTCTAGCTGCGGAAGCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1238  
QY 1248 GAAGATTACTGACCTGGGAAGCATCAGCGCGGAGGGGCTGAGCGGGTGATCAGCGTGA 1307  
Db 1239 GAAGATTACTGACCTGGGAAGCATCAGCGCGGAGGGGCTGAGCGGGTGATCAGCGTGA 1298  
QY 1308 GATGGAGATCCCGGCTTCCATGCGCTCTCATCCAGGAAATGTTGGAACTCAGAGG 1367  
Db 1299 GATGGAGATCCCGGCTTCCATGCGCTCTCATCCAGGAAATGTTGGAACTCAGAGG 1358  
QY 1368 CTGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGAGCGGGTGGCTGCGCC 1427  
Db 1359 CTGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGAGCGGGTGGCTGCGCC 1418  
QY 1428 CCGCGCAGGAGCTGTAGCCCGCAGCTCAGCCCGAGCTCCACAGAGAGCGCGCGCAC 1487  
Db 1419 CCGCGCAGGAGCTGTAGCCCGCAGCTCAGCCCGAGCTCCACAGAGAGCGCGCGCAC 1478  
QY 1488 CCACCTCCCTGACCGCCACCGCCACATGACACAGCCCTGCGCTCCGCCCCGGCTTTT 1547  
Db 1479 CCACCTCCCTGACCGCCACCGCCACATGACACAGCCCTGCGCTCCGCCCCGGCTTTT 1538  
QY 1548 CTCTGCTTCTTACCGACCATGTGACCCCGCAGCCCGCTGCGCCACCTGCGCTCCCGG 1607  
Db 1539 CTCTGCTTCTTACCGACCATGTGACCCCGCAGCCCGCTGCGCCACCTGCGCTCCCGG 1598  
QY 1608 GCAGTACTGGGACCTTCTCTGGGGGACGGGAGGAGGAGCGGAGCTCTTGGACAG 1667  
Db 1599 GCAGTACTGGGACCTTCTCTGGGGGACGGGAGGAGGAGGAGCGGAGCTCTTGGACAG 1658  
QY 1668 AGGCTTGGGCGCTCAGTGGAGCTGCTCTCCACAGCCTGGGTGAGCTCAGAGGCGGAG 1727  
Db 1659 AGGCTTGGGCGCTCAGTGGAGCTGCTCTCCACAGCCTGGGTGAGCTCAGAGGCGGAG 1718  
QY 1728 GCCAGGNACTGAGTGGAGCGGCTTCTGGTCTCAGGATGGGTCTCGGGGCGCTCGTG 1787

Db 1719 GCCAGAACTGAGTGAGGCCCTGTGCTCGGTCTCAGGATGGTCTCGGGGCTCGTG 1778  
QY 1788 TTCTATCAAGACACCCCTCTGCCCCAGCTCACCACATCTTTCATCACCAGCAAAAGCCAGGAC 1847  
Db 1779 TTCTATCAAGACACCCCTCTGCCCCAGCTCACCACATCTTTCATCACCAGCAAAAGCCAGGAC 1838  
QY 1848 TTGGTCCCCATCTCTCAGAACTCAGAGCCATTCGTCGCCAGCTGGGGAACCTCAACCT 1907  
Db 1839 TTGGTCCCCATCTCTCAGAACTCAGAGCCATTCGTCGCCAGCTGGGGAACCTCAACCT 1898  
QY 1908 CCCCCCTCCCTCGGTGTTGTGACAGAGGGGTGGGACAGGGGGTTCGCCCTGTAC 1967  
Db 1899 CCCCCCTCCCTCGGTGTTGTGACAGAGGGGTGGGACAGGGGGTTCGCCCTGTAC 1958  
QY 1968 ATACCTGCCATACCAACCCAGGTATTATCTCGCTGGTGTCTTTTATTTTAAATTT 2027  
Db 1959 ATACCTGCCATACCAACCCAGGTATTATCTCGCTGGTGTCTTTTATTTTAAATTT 2018  
QY 2028 TTTGTTTTGATTTTTTAAAGAAATTTTCATTTTAAAGCAAAAATAAAAAA 2085  
Db 2019 TTTGTTTTGATTTTTTAAAGAAATTTTCATTTTAAAGCAATTTATCTGAAGAA 2076

## RESULT 14

AAS83047  
ID AAS83047 standard; cDNA; 3052 BP.

XX AAS83047;

AC AC  
XX XX  
DT 13-FEB-2002 (first entry)

XX XX  
DE DNA encoding novel human diagnostic protein #18851.

XX XX  
KW Human; chromosome mapping; gene mapping; gene therapy; forensic;  
KW food supplement; medical imaging; diagnostic; genetic disorder; ss.

OS Homo sapiens.

XX WO200175067-A2.

PN 11-OCT-2001.

XX 30-MAR-2001; 2001WO-US08631.

PF 31-MAR-2000; 2000US-0540217.

PR 23-AUG-2000; 2000US-0649167.

XX (HYSE-) HYSEQ INC.

XX Drmanac RT, Liu C, Tang YT;

XX WPI; 2001-639362/73.

DR P-PSDB; ABG18860.

XX XX  
PT New isolated polynucleotide and encoded polypeptides, useful in  
PT diagnostics, forensics, gene mapping, identification of mutations  
PT responsible for genetic disorders or other traits and to assess  
PT biodiversity

XX Claim 1; SEQ ID No 18851; 103pp; English.

XX CC  
CC The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating



disorders involving aberrant protein expression or biological activity.  
The polypeptide and polynucleotide sequences have applications in  
diagnostics, forensics, gene mapping, identification of mutations  
responsible for genetic disorders or other traits to assess biodiversity  
and to produce other types of data and products dependent on DNA and  
amino acid sequences. AAS64197-AAS94564 represent novel human  
diagnostic coding sequences of the invention.  
Note: The sequence data for this patent did not appear in the printed  
specification, but was obtained in electronic format directly from WIPO  
at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 3052 BP; 613 A; 1020 C; 894 G; 525 T; 0 other;

Query Match 81.9%; Score 1707.4; DB 23; Length 3052;  
Best Local Similarity 99.2%; Pred. No. 2.2e-309;  
Matches 1789; Conservative 0; Mismatches 6; Indels 8; Gaps 7;

```
QY 290 CCATTGAGACCCAGAGCAGCTTCTGAAGAGATAGTCCCGACCCCTCCCTCGCCACCCC 349
D 1250 CCATTGAGACCCAGAGCAGCTTCTGAAGAGATAGTCCCGACCCCTCCCTCGCCACCCC 1309
QY 350 CTCATCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAGTCTCAGGCTACCACT 409
D 1310 CTCATCCCGCATCTACAAGCCTTGTCTGTCTGTCAGGACAAGTCTCAGGCTACCACT 1369
QY 410 ATGGGGTCAGCGCTGTGAGGGCTGCAAGGGCTTCTCCGCGCAGCATCAAGTCAAGAGAACA 469
D 1370 ATGGGGTCAGCGCTGTGAGGGCTGCAAGGGCTTCTCCGCGCAGCATCAAGTCAAGAGAACA 1429
QY 470 TGGTGTACACCTGTCACGGGACAGAACTGCATCATCAAGAGTGCACCGGAAACCGCT 529
D 1430 TGGTGTACACCTGTCACGGGACAGAACTGCATCATCAAGAGTGCACCGGAAACCGCT 1489
QY 530 GCCAGTACTGCCGACTGCAGAGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAGAA 589
D 1490 GCCAGTACTGCCGACTGCAGAGTCTTGAAGTGGGATGTCCAAAGAGTCTGTGAGAA 1549
QY 590 ACGACCAAGCAAGAAAGAGAGTGCACCAAGCGCGAGTCTCTGAGAGCTACACGC 649
D 1550 ACGACCAAGCAAGAAAGAGAGTGCACCAAGCGCGAGTCTCTGAGAGCTACACGC 1609
QY 650 TGACCGCGAGGTGGGGAGCTCATTTGAGAGTGGCGCAAGCGCGAGTCTCTGAGAGCTACACGC 709
D 1610 TGACCGCGAGGTGGGGAGCTCATTTGAGAGTGGCGCAAGCGCGAGTCTCTGAGAGCTACACGC 1669
QY 710 CTGCGCTCTGCCAGCTGGGCAATACACTAGCAACAGCTCAGCAACAGCTGATCTCTC 769
D 1670 CTGCGCTCTGCCAGCTGGGCAATACACTAGCAACAGCTCAGCAACAGCTGATCTCTC 1729
QY 770 TGGACATGACCTCTGGGACAGTTTCACTGATCTCCACCAAGTGCATCAATTAAGACTG 829
D 1730 TGGACATGACCTCTGGGACAGTTTCACTGATCTCCACCAAGTGCATCAATTAAGACTG 1789
QY 830 TGAGATTTCGCAAGCAGCTCCCGGCTTCAACACCTTCAACCTGCGGATCTGACGCGGTACACGCC 889
D 1790 TGAGATTTCGCAAGCAGCTCCCGGCTTCAACACCTTCAACCTGCGGATCTGACGCGGTACACGCC 1849
QY 890 TCCTCAAGGCTGCTGCTGGACATCTGATCTCTGCGGATCTGACGCGGTACACGCC 949
D 1850 TCCTCAAGGCTGCTGCTGGACATCTGATCTCTGCGGATCTGACGCGGTACACGCC 1909
QY 950 AGCAGGACACCATGACCTTCTCGGACGGGTGACCCCTGAAACCGGACCGAGATGACAAAG 1009
D 1910 AGCAGGACACCATGACCTTCTCGGACGGGTGACCCCTGAAACCGGACCGAGATGACAAAG 1969
QY 1010 CTGCTTTCGCGCCCTTCAACGAGCTGCTTTCGCTTTCGCGCAACGAGTCTGCTCCCTGG 1069
D 1970 CTGCTTTCGCGCCCTTCAACGAGCTGCTTTCGCTTTCGCGCAACGAGTCTGCTCCCTGG 2029
QY 1070 AGATGATGATGGGAGACGGGCTGCTCAGCGCATCTGCTCCTCATCTGCGGAGACCGCC 1129
D 2030 AGATGATGATGGGAGACGGGCTGCTCAGCGCATCTGCTCCTCATCTGCGGAGACCGCC 2089
```

RESULT 15  
AAS83048  
ID AAS83048 standard; cdna; 2930 BP.  
XX

```
QY 1130 AGGACCTGAGACGACCGGACCGGCTGACATGCTGCAGGAGCGCGCTCTCTGAGAGCGGTAA 1189
D 2090 AGGACCTGAGACGACCGGACCGGCTGACATGCTGCAGGAGCGCGCTCTCTGAGAGCGGTAA 2149
QY 1190 AGGTCTACCTGCGGAAGCGGAGCGGACCGGCGCCACATGTTCCCAAGATGCTAATGA 1249
D 2150 AGGTCTACCTGCGGAAGCGGAGCGGACCGGCGCCACATGTTCCCAAGATGCTAATGA 2209
QY 1250 AGATTACTACCTGCGGAAGCATCAGCGCCAAAGGGGCTGAGCGGGTGCATCAGCTGAAGA 1309
D 2210 AGATTACTACCTGCGGAAGCATCAGCGCCAAAGGGGCTGAGCGGGTGCATCAGCTGAAGA 2269
QY 1310 TGGAGATCCCGGCTCCATGCGCGCTTCTCATCAGGAAATGTTGGAGAACTCAGAGGGCC 1369
D 2270 TGGAGATCCCGGCTCCATGCGCGCTTCTCATCAGGAAATGTTGGAGAACTCAGAGGGCC 2329
QY 1370 TGGACACTCTGAGCGGACAGCGGGGGTGGGGGGGCGGACGGGGTGGCGCTGCCCGCCC 1429
D 2330 TGGACACTCTGAGCGGACAGCGGGGGTGGGGGGGCGGACGGGGTGGCGCTGCCCGCCC 2389
QY 1430 CCGCAGGACAGCTGTAGCCCGCAGCCTCAGCCCGCAGCTTCCAAACAGAGCAGCGCGCCACCC 1489
D 2390 CCGCAGGACAGCTGTAGCCCGCAGCCTCAGCCCGCAGCTTCCAAACAGAGCAGCGCGCCACCC 2449
QY 1490 ACTCCCGCTGACCGCCCGCAGCAGCATGACAGCCCTCGCCCTCGCCCGCCCGCCT -TTTTT 1548
D 2450 ACTCCCGCTGACCGCCCGCAGCAGCATGACAGCCCTCGCCCTCGCCCGCCCGCCTTTTTC 2509
QY 1549 TCTGCTTCTTACCGACCATGTGACCCC -GCACACGCGCTGCCCGCCCGCCT -GCCCTTCC 1605
D 2510 TCTGCTTCTTACCGACCATGTGACCCC -GCACACGCGCTGCCCGCCCGCCTTTTTCCTCG 2569
QY 1606 GGGCAGTACTGGGGACCTTCCCTGGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1665
D 2570 GGGCAGTACTGGGGACCTTCCCTGGGGGAGCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2629
QY 1666 AGAGGCTTGGGCGCTCAGTGGAGTGCCTGCTCCACAGCGCTGGGCTGAGCTCAGAGGCGCG 1725
D 2630 AGAGGCTTGGGCGCTCAGTGGAGTGCCTGCTCCACAGCGCTGGGCTGAGCTCAGAGGCGCG 2689
QY 1726 AGG -CCAGGAACTGAGTGGGCGCTGGT -CCTGGGTCTCAGAGTGGGT -CCTGGGGGCG 1782
D 2690 AGGCGCGGAACTGAGTGGGCGCTGGTGGGCTCTCAGGATGGGCTCTGGGCGCTGGGCGCG 2749
QY 1783 TCGTCTTCAATCAAGACACCGCTCTGCCAGCTCAGGAGTGGGCTCTCAGGATGGGCTCTGGGCGCG 1842
D 2750 TCGTCTTCAATCAAGACACCGCTCTGCCAGCTCAGGAGTGGGCTCTCAGGATGGGCTCTGGGCGCG 2809
QY 1843 AGGACTTGGCTCCCGCTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACT 1901
D 2810 AGGACTTGGCTCCCGCTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACT 2869
QY 1902 CAACCTCCCGCTGCGCTGGTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1961
D 2870 CAACCTCCCGCTGCGCTGGTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2929
QY 1962 CTGTACATACCTGCGCTGCGCTGGTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2021
D 2930 CTGTACATACCTGCGCTGCGCTGGTGGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2989
QY 2022 TAAATTTTGTGTTGTTGTTTAAATGAAATTTTAAATGAAATTTTAAATGAAATTTTAAATGAAATTTT 2081
D 2990 TAAATTTTGTGTTGTTGTTTAAATGAAATTTTAAATGAAATTTTAAATGAAATTTTAAATGAAATTTT 3049
QY 2082 AAA 2084
D 3050 AAA 3052
```

AAS83048;  
 13-FEB-2002 (first entry)  
 DNA encoding novel human diagnostic protein #18852.  
 Human; chromosome mapping; gene mapping; gene therapy; forensic;  
 food supplement; medical imaging; diagnostic; genetic disorder; ss.  
 Homo sapiens.  
 WO200175067-A2.  
 11-OCT-2001.  
 30-MAR-2001; 2001WO-US08631.  
 31-MAR-2000; 2000US-0540217.  
 23-AUG-2000; 2000US-0649167.  
 (HYSB-) HYSEQ INC.  
 Drmanac RT, Liu C, Tang YT;  
 WPI; 2001-639362/73.  
 P-PSDB; ABG18861.  
 New isolated polynucleotide and encoded polypeptides, useful in  
 diagnostics, forensics, gene mapping, identification of mutations  
 responsible for genetic disorders or other traits and to assess  
 biodiversity -  
 Claim 1; SEQ ID NO 18852; 103pp; English.  
 The invention relates to isolated polynucleotide (I) and  
 polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 and gene mapping, and in recombinant production of (II). The  
 polynucleotides are also used in diagnostics as expressed sequence tags  
 for identifying expressed genes. (I) is useful in gene therapy techniques  
 to restore normal activity of (II) or to treat disease states involving  
 (II). (II) is useful for generating antibodies against it, detecting or  
 quantitating a polypeptide in tissue, as molecular weight markers and as  
 a food supplement. (II) and its binding partners are useful in medical  
 imaging of sites expressing (II). (I) and (II) are useful for treating  
 disorders involving aberrant protein expression or biological activity.  
 The polypeptide and polynucleotide sequences have applications in  
 diagnostics, forensics, gene mapping, identification of mutations  
 responsible for genetic disorders or other traits to assess biodiversity  
 and to produce other types of data and products dependent on DNA and  
 amino acid sequences. AAS64197-AAS94564 represent novel human  
 diagnostic coding sequences of the invention.  
 Note: the sequence data for this patent did not appear in the printed  
 specification, but was obtained in electronic format directly from WIPO  
 at ftp.wipo.int/pub/published\_pct\_sequences.  
 Sequence 2930 BP; 565 A; 1008 C; 807 G; 550 T; 0 other;  
 Query Match 73.8%; Score 1540.2; DB 23; Length 2930;  
 Best Local Similarity 97.2%; Pred. No. 3.5e-278;  
 Matches 1770; Conservative 0; Mismatches 28; Indels 23; Gaps 19;  
 288 CTTCCATTGAGACCCAGAGCAGAGTTCTGAAGAGATAGTGCACGCCCTCCCTCGCCACC 347  
 279 CACCATTGAGACCCAGAGCAGAGTTCTGAAGAGATAGTGCACGCCCTCCCTCGCCACC 338  
 348 CCTCTACCCCGCATCTACAAGCTTGTCTGCTCAGGACAAGTCTCTCAGGCTACCA 407  
 339 CCTCTACCCCGCATCTACAAGCTTGTCTGCTCAGGACAAGTCTCTCAGGCTACCA 398  
 408 CTATGGGCTCAGCGCTGTGAGGGCTGCAAGGCTTCTTCCGCCGCGCAGCATCCAGAGAA 467  
 399 CTATGGGCTCAGCGCTGTGAGGGCTGCAAGGCTTCTTCCGCCGCGCAGCATCCAGAGAA 458

QY 468 CATGTGTACAGTGTGTACACGGGACAAGAACTGCATCATCAACAAGGTGACCCGGAAACC 527  
 DB 459 CATGTGTACAGTGTGTACACGGGACAAGAACTGCATCATCAACAAGGTGACCCGGAAACC 518  
 QY 528 CTGCCAGTACTGCCGACTGCAGAACTGCTTTGAAGTGGGCATGTCCAAGGAGTCTGTGAG 587  
 DB 519 CTGCCAGTACTGCCGACTGCAGAACTGCTTTGAAGTGGGCATGTCCAAGGAGTCTGTGAG 578  
 QY 588 AAACGACCGAAACAAGAAGAAGAGAGGTGCCCAAGCCCGAGTCTCTGAGAGCTACAC 647  
 DB 579 AAACGACCGAAACAAGAAGAAGAGAGGTGCCCAAGCCCGAGTCTCTGAGAGCTACAC 638  
 QY 648 GCTGACCGCGGAGGTGGGGAGCTCATTTGAGAAGTGGCGCAAGCGCACCA-GGAACCT 706  
 DB 639 GCTGACCGCGGAGGTGGGGAGCTCATTTGAGAAGTGGCGCAAGCGCACCAAGGAAACCT 698  
 QY 707 TCCCTGCCCTCTGCCAGCTGGGCAAAATACACTAC-GAACAACAGC-TCAAGAACAAGCTG- 763  
 DB 699 TCCCTGCCCTCTGCCAGCTGGGCAAAATACACTACAGCAACAGCTTCAACAACAGCTGT 758  
 QY 764 TCTCTCTGGACATTGACCTCTGGGACAAGTTCACTGAGTGAATCTTCCACCAAGTGCATTA 823  
 DB 759 TCTCTCTGGACATTGACCTCTGGGACAAGTTCACTGAGTGAATCTTCCACCAAGTGCATTA 818  
 QY 824 AGACTGTGG-AGTTGCCAAG--CAGCTGCCGGCTTCAACACCTCACCATC-GCCGAC 879  
 DB 819 AAAATGGGAGGTGGCGCAAGGCTGCGGGCTTCAACACCTCACCATCAGTCGCGGAC 878  
 QY 880 CAGATACCCCTCCCAA-GGCTGCTGCT--GGACATCTGATCTCTGCGGATCTGCACGC 937  
 DB 879 CAGATACCCCTCCCAAAGGGCTGCTGCTGGGACATCTGATCTGGGATCTGCACGC 938  
 QY 938 GGTACACGCCGAG--CAGGACACCATGACCTTC-TGCGAGGGGTGACCTTGAACC-GG 993  
 DB 939 GGTACACGCCGAGGACGACCATGACCTTCTTGGAGGGGTGACCTTGAACC-GG 998  
 QY 994 ACCCAGATGCAACAGCTGG--CTTGGCCCTCTACCGACCTGTCTTTGGC-TTGGCC 1050  
 DB 999 ACCCAGATGCAACAGCTGGGCTTGGCCCTCTTACCGACCTGTCTTTGGCTTTGGCC 1058  
 QY 1051 AACCAGTGTCTGCCCT--GGAGATGGATGATG--CGGAGAGGGGTGCTCAGCGCCATC. 1107  
 DB 1059 AACCAGTGTCTGCCCTTGGAGATGGATGATGGGAGAGAGGGGTGCTCAGCGCCATC 1118  
 QY 1108 TGCTCATCTCGGAGACCGCCAGGACCTGGAGCAGCGGACCGGGTGGATGCTGAG 1167  
 DB 1119 TGCTCATCTCGGAGACCGCCAGGACCTGGAGCAGCGGACCGGGTGGATGCTGAG 1178  
 QY 1168 GAGCCGTGTGGAGGCGCTAAAGTCTTACGTGCGGAAGCGGAGCGCCAGCCGCCAC 1227  
 DB 1179 GAGCCGTGTGGAGGCGCTAAAGTCTTACGTGCGGAAGCGGAGCGCCAGCCGCCAC 1238  
 QY 1228 ATGTTCCCAAGATGCTAATCAAGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCT 1287  
 DB 1239 ATGTTCCCAAGATGCTAATCAAGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCT 1298  
 QY 1288 GAGCGGTGATCAGCTGAAGATGGAGATCCCGGGCTTCCATGCCGCTCTCATCCAGAA 1347  
 DB 1299 GAGCGGTGATCAGCTGAAGATGGAGATCCCGGGCTTCCATGCCGCTCTCATCCAGAA 1358  
 QY 1348 ATTTTGAGAACTCAGAGGCGCTGGACACTCTTGAGCGGACAGCGGGGGTGGGGCGG 1407  
 DB 1359 ATTTTGAGAACTCAGAGGCGCTGGACACTCTTGAGCGGACAGCGGGGGTGGGGCGG 1418  
 QY 1408 GAGCGGGTGGCTGCCCGCCGCCA-GGAGCTGTAGCCCCAGCTCAGCCCCAGCTC 1466  
 DB 1419 GAGCGGGTGGCTGCCCGCCGCCAAGCTGTAGCCCCAGCTCAGCCCCAGCTC 1478  
 QY 1467 CAACAGAAGCAGCGCGCCACCTCCCGTGTACCGCCGACGACCATGACACAGGCC 1526  
 DB 1479 CAACAGAAGCAGCGCGCCACCTCCCGTGTACCGCCGACGACCATGACACAGGCC 1538

QY 1527 TCGCCCTCCGCCCGGGCTTTCTCTGCTTTCTACCGACATGTGACCCCGCAGCC 1586  
 Db 1539 TCGCCCTCCGCCCGGGCTTTCTCTGCTTTCTACCGACATGTGACCCCGCAGCC 1598  
 QY 1587 TCGCCCCACCTGCGCT-CCGCGCAGTACTGGGACCTTCCCTGGGGACGGGAGGAG 1645  
 Db 1599 TCGCCCCACCTGCGCTGCGCGGCGAGTACTGGGACCTTCCCTGGGGACGGGAGGAG 1658  
 QY 1646 GAGGCAGGACTCCTTGGACAGAGGCTGGGCCCTCAGTGGACTGCTGCCACAGCC 1705  
 Db 1659 GAGGCAGGACTCCTTGGACAGAGGCTGGGCCCTCAGTGGACTGCTGCCACAGCC 1718  
 QY 1706 TGGGCTGAGCTCAGAGCCGAGCCAGGAACCTGAGTGAGGCCCTGCTGGGTCTCAG 1765  
 Db 1719 TGGGCTGAGCTCAGAGCCGAGCCAGGAACCTGAGTGAGGCCCTGCTGGGTCTCAG 1778  
 QY 1766 GATGGTCTCGGGGCTCGTGTTCATCAGACACCCCTCTGCCAGCTCACCACATCTT 1825  
 Db 1779 GATGGTCTCGGGGCTCGTGTTCATCAGACACCCCTCTGCCAGCTCACCACATCTT 1838  
 QY 1826 CATCACGACAAAGCCAGGACTTGGTCCGCCATCCTC-AGACTCACAAGCCATTGCT 1884  
 Db 1839 CATCACGACAAAGCCAGGACTTGGTCCGCCATCCTCAGAACTCACAAGCCATTGCT 1898  
 QY 1885 CCCCAGCTGGGAACTCAACCTCCGCCCTGCGTGGTGGTGGAGAGGGGTGGGACA 1944  
 Db 1899 CCCCAGCTGGGAACTCAACCTCCGCCCTGCGTGGTGGTGGAGAGGGGTGGGACA 1958  
 QY 1945 GGGCGGGGGTTCGCCCTGTACATACCCCTGCCATACCAACCCAGGTATTATTCGCG 2004  
 Db 1959 GGGCGGGGGTTCGCCCTGTACATACCCCTGCCATACCAACCCAGGTATTATTCGCG 2018  
 QY 2005 TGGTTTGTATTTATTTATTTATTTTGTGTTTGTATTTTATTAAGAAATTTTCAATTTA 2064  
 Db 2019 TGGTTTGTATTTATTTATTTATTTTGTGTTTGTATTTTATTAAGAAATTTTCAATTTA 2078  
 QY 2065 AGCACAAAAA 2085  
 Db 2079 AGCACATTTATCTGAAGAA 2099

Search completed: March 29, 2003, 13:27:12  
 Job time : 447 secs

pred No. is the number of results predicted by chance to have a

RESULT 1  
 HSM805408  
 Locus  
 DEFINITION  
 Accession  
 AL834159  
 Version  
 AL834159.1  
 GI:21739638  
 Keywords  
 human.  
 Source  
 Homo sapiens  
 Organism  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 2300)  
 Reference  
 Ansong.W., Winkner,U., Mewes,H.W., Weil,B. and Wiemann,S.  
 Authors  
 Direct Submission  
 Title  
 Submitted (09-JUL-2002) 1, D-85764 Neuherberg, GERMANY  
 Journal

COMMENT Clone from S. Wiemann, Molecular Genome Analysis, German Cancer Research Center (DFPZ); Email s.wiemann@dkfz-heidelberg.de; sequenced by EMBL (European Molecular Biology Laboratories, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.  
This clone (DKFZp761C0417) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de Further information about the clone and the sequencing project is available at <http://mips.gsf.de/proj/cDNA/>.

## FEATURES

## source

1. 2300  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="DKFZp761C0417"  
/tissue\_type="amygdala"  
/clone\_lib="761 (synonym: hamy2). Vector pSport1; host DH10B; sites NotI + SalI"  
/dev\_stage="adult"

polyA\_site 462 a 765 c 656 g 417 t  
BASE COUNT  
ORIGIN

Query Match 95.6%; Score 1995; DB 9; Length 2300;  
Best Local Similarity 99.2%; Pred. No. 0;

Matches 2058; Conservative 0; Mismatches 10; Indels 7; Gaps 5;

QY 11 GGGCGAGGGGACGCTCTCTCCGCCAGCTGCTCGTTCGGATGGCGCGCGCGC--TGA 68  
DB 228 GGGCGAGGGGACGCTCTCTCTCCGCCAGCTGCTGTCGGATGGCGCGCGCGCGCTGAG 287  
QY 69 GTGACGGGGGGCGGCGCGACGACTTCCAGCTCGGACCTCTT--GCCTTCGAGGGGAAAG 126  
DB 288 TTGACGGGGGGCGGCGCGACGACTTCCAGCTCGGACCTCTTGGCTTTGAGGGGAAAG 347  
QY 127 ATGTACGAGAGTGTAGAGTGGGGGTCCACACCCCTAATCCCTTCCTAGTGGTGGATTT 186  
DB 348 ATGTACGAGAGTGTAGAGTGGGGGTCCACACCCCTAATCCCTTCCTAGTGGTGGATTT 407  
QY 187 TATTAACAGAACCGGGCTCTTTGCTCCAGAGAGGGGTCCCGCGCGCGGGTCCCTAC 246  
DB 408 TATTAACAGAACCGGGCTCTTTGCTCCAGAGAGGGGTCCCGCGCGGGTCCCTAC 467  
QY 247 TCACACCGCGCTCGGACTCGCGCTTTGGAATGGCTCAACACCTCCATTGAGACCCAGAGC 306  
DB 468 TCACACCGCGCTCGGACTCGCGCTTTGGAATGGCTCAACACCTCCATTGAGACCCAGAGC 527  
QY 307 AGCAGTCTGAAGAGATAGTGGCCAGCCCTCCCTCGCCACCCCTCTACCCCGCATCTAC 366  
DB 528 AGCAGTCTGAAGAGATAGTGGCCAGCCCTCCCTCGCCACCCCTCTACCCCGCATCTAC 587  
QY 367 AAGCCTTGCTTCTGTCTAGGACAGTCTCTCAGGCTACCACTATGGGTCAGGCGCTGT 426  
DB 588 AAGCCTTGCTTGTCTGTAGGACAGTCTCTCAGGCTACCACTATGGGTCAGGCGCTGT 647  
QY 427 GAGGGTGAAGGGCTCTTCCGCCAGCATCCAGAGAACATGGTGTACAGTGTCTAC 486  
DB 648 GAGGGTGAAGGGCTCTTCCGCCAGCATCCAGAGAACATGGTGTACAGTGTCTAC 707  
QY 487 CGGGACAAGAACTGCAATCAACAAGGTGACCCGGAACCCCTGCCAGTACTGCCAGCTG 546  
DB 708 CGGGACAAGAACTGCAATCAACAAGGTGACCCGGAACCCCTGCCAGTACTGCCAGCTG 767  
QY 547 CAGAAGTGTCTTCAAGTGGGCATGTCCAGAGGATCTGTGAGAAACGACGAAACAAGAAG 606  
DB 768 CAGAAGTGTCTTCAAGTGGGCATGTCCAGAGGATCTGTGAGAAACGACGAAACAAGAAG 827  
QY 607 AAGAAGGAGGTGCCAAGCCCGAGTCTCTGAGAGCTACACGCTGACGCGGAGGTGGGG 666  
DB 828 AAGAAGGAGGTGCCAAGCCCGAGTCTCTGAGAGCTACACGCTGACGCGGAGGTGGGG 887  
QY 667 GAGCTATTGAGAGGTGGCAAGCGCACACGAAACCTTCCTGCCCTCTGCCAGCTG 726  
|||||

DB 888 GAGCTCATTTAGAA--GTGGCAAAAGCGCACAGGAAACCTTCCTGCTCTGCCAGCTG 946  
QY 727 GCAAAATACACTACGAACAACAGCTCAGACACAGTGTCTCTCTGACATTGACCTCTGG 786  
DB 947 GCAAAATACACTACGAACAACAGCTCAGAACACAGTGTCTCTCTGACATTGACCTCTGG 1006  
QY 787 GACAAAGTTCACTGAGTAACTCTCC--ACCAAGTGCATCATTAAGAGCTGTGGAGTTCGCCAAGCA 845  
DB 1007 GACAAAGTTCACTGAGTAACTCTCCAAACCAAGTGCATCATTAAGAGCTGTGGAGTTCGCCAAGCA 1066  
QY 846 GCTGCCCGGGCTTCAACACCTCACCATCGCCGACAGATCACCTCTCAAGGCTGCCTG 905  
DB 1067 GCTGCCCGGGCTTCAACACCTCACCATCGCCGACAGATCACCTCTCAAGGCTGCCTG 1126  
QY 906 CTTGACATCTCTGATCTCTGCGGATCTGCACGCGGTACACGCCGAGCAGACACCATGAC 965  
DB 1127 CTTGACATCTCTGATCTCTGCGGATCTGCACGCGGTACACGCCGAGCAGACACCATGAC 1186  
QY 966 CTTCTCGGACGGGCTGACCTTGAACCGGACCCAGATGCACAAACGGTGGCTTCGGGCCCTT 1025  
DB 1187 CTTCTCGGACGGGCTGACCTTGAACCGGACCCAGATGCACAAACGGTGGCTTCGGGCCCTT 1246  
QY 1026 CACGACCTGTGCTTTGCTTCGCCCAACAGATGCTGCCCTTGGAGATGGATGATCGGA 1085  
DB 1247 CACGACCTGTGCTTTGCTTCGCCCAACAGATGCTGCCCTTGGAGATGGATGATCGGA 1306  
QY 1086 GACGGGGTGTCTACGCGCATCTGCCTCATCTGCGGAGAGCCGAGGACCTGGAGGAGCC 1145  
DB 1307 GACGGGGTGTCTACGCGCATCTGCCTCATCTGCGGAGAGCCGAGGACCTGGAGGAGCC 1366  
QY 1146 GGACCGGTGTGACATGCTGTCAGGAGCGCTGCTGTCAGGCGCTTAAAGGTCTACGTCGGAA 1205  
DB 1367 GGACCGGTGTGACATGCTGTCAGGAGCGCTGCTGTCAGGCGCTTAAAGGTCTACGTCGGAA 1426  
QY 1206 GCGAGGCGCCAGCGCCGCCACATGTTCCCAAGATGCTAATGAAGTACTGACTGCG 1265  
DB 1427 GCGAGGCGCCAGCGCCGCCACATGTTCCCAAGATGCTAATGAAGTACTGACTGCG 1486  
QY 1266 AAGCATCAGCGCCCAAGGGGCTGAGCGGGTGTACGCTGAAGATGGAGATCCCGGGCTC 1325  
DB 1487 AAGCATCAGCGCCCAAGGGGCTGAGCGGGTGTACGCTGAAGATGGAGATCCCGGGCTC 1546  
QY 1326 CATCCCGCTCTCATCCAGGAAATGTTGAGAACTTCAGAGGCGCTTGACACTCTCAGCGG 1385  
DB 1547 CATCCCGCTCTCATCCAGGAAATGTTGAGAACTTCAGAGGCGCTTGACACTCTCAGCGG 1606  
QY 1386 ACAGCCGGGGGTGGGGGGGAGCGGGGTGGCTGCCCCCGCCGAGGAGGAGTGTAG 1445  
DB 1607 ACAGCCGGGGGTGGGGGGGAGCGGGGTGGCTG--GCCCGCCGAGGAGGAGTGTAG 1665  
QY 1446 CCCAGAGCTCAGCCCCAGCTCCAAAGAGCAGCCGCGCCACCCACTCCCCGTGACGCC 1505  
DB 1666 CCCAGAGCTCAGCCCCAGCTCCAAAGAGCAGCCGCGCCACCCACTCCCCGTGACGCC 1725  
QY 1506 CAGCCACATGGACACAGCCCTCGCCCTCGCCCGCGGGCTTTTCTGCTGCTTTCTACCGAC 1565  
DB 1726 CAGCCACATGGACACAGCCCTCGCCCTCGCCCGCGGGCTTTTCTGCTGCTTTCTACCGAC 1785  
QY 1566 CATGTGACCCCGACACAGCCCTGCCCCACCTGCGCTCCCGGGCAGTACTGGGAGCTTC 1625  
DB 1786 CATGTGACCCCGACACAGCCCTGCCCCACCTGCGCTCCCGGGCAGTACTGGGAGCTTC 1845  
QY 1626 CTTGGGGAGCGGGAGGAGGAGGAGCGACTCTCTTGGACAGAGGCTTGGGCCCTCTAGTG 1685  
DB 1846 CTTGGGGAGCGGGAGGAGGAGGAGGAGCGACTCTCTTGGACAGAGGCTTGGGCCCTCTAGTG 1905  
QY 1686 GACTGCTGTCTCCACAGCCCTGGCGTACOTCAGAGGCGGAGGAGGAGGAGTGTAGTGAGG 1745  
DB 1906 GACTGCTGTCTCCACAGCCCTGGCGTACOTCAGAGGCGGAGGAGGAGGAGTGTAGTGAGG 1965  
QY 1746 CCCTGCTGTGCTGCTCAGGATGGGTTCCTGGGGCTCTGTTTCATCAAGACACCCCTC 1805  
DB 1966 CCCTGCTGTGCTGCTCAGGATGGGTTCCTGGGGCTCTGTTTCATCAAGACACCCCTC 2025  
|||||

```
QY 1806 TGCCAGCTCACCACATCTTCATCACCAGCAAGCCGAGGACTGGCTCCCCCATCTCA 1865
|
|
|
Db 2026 TGCCAGCTCACCACATCTTCATCACCAGCAAGCCGAGGACTGGCTCCCCCATCTCA 2085
|
|
|
QY 1866 GAACACTACAAGCCATGTCTCCAGCTGGGGAGACCTCAACCTCCCCCTGCTCGGTGG 1925
|
|
|
Db 2086 GAACACTACAAGCCATGTCTCCAGCTGGGGAGACCTCAACCTCCCCCTGCTCGGTGG 2145
|
|
|
QY 1926 TGACAGAGGGGTGGGACAGGGGGGGGGTTCCTCCCTGTACATACCTGCCATACCAAC 1985
|
|
|
Db 2146 TGACAGAGGGGTGGGACAGGGGGGGGGTTCCTCCCTGTACATACCTGCCATACCAAC 2205
|
|
|
QY 1986 CCAGGTATTAATCTCGCTGGTTCCTTTTATTTTAAATTTTGTGTTGATTTTTT 2045
|
|
|
Db 2206 CCAGGTATTAATCTCGCTGGTTCCTTTTATTTTAAATTTTGTGTTGATTTTTT 2265
|
|
|
QY 2046 AATAAGATTTTCATTTTAAGCAACAAAAA 2080
|
|
|
Db 2266 AATAAGATTTTCATTTTAAGCAACAAAAA 2300
|
|
|

RESULT 2
BC008727 2441 bp mRNA linear PRI 12-JUL-2001
LOCUS Homo sapiens, retinoic acid receptor, alpha, clone MGC:1651
DEFINITION IMAGE:3163891, mRNA, complete cds.
ACCESSION BC008727
VERSION BC008727.1 GI:14250549
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 2441)
Direct Submission
Submitted (25-MAY-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
NIH-MGC Project URL: http://mgc.nci.nih.gov
Contact: MGC help desk
Email: cgabbs@email.nih.gov
Tissue Procurement: AFCC
cDNA Library Preparation: Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
Contact: nisc.mgc@nih.gov
Shenchenko, Y., Wetherby, K.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, O.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Snyder, B., Stantripop, S., Thomas, P.J.,
Tongson, E.E., Touchman, J.W., Tsurgon, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAL Plate: 5 Row: 9 Column: 23
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 4506418.
FEATURES
Location/Qualifiers
1..2441
/organism="Homo sapiens"
/db_xref="LocusID:5914"
/db_xref="taxon:9606"
/clone="MGC:1651 IMAGE:3163891"
/tissue_type="Brain, neuroblastoma"
/clone_lib="NIH_MGC_19"
```

```
/lab_host="DH10B-R"
/notes="Vector: pOTB7"
465..1853
/codon_start=1
/product="retinoic acid receptor, alpha"
/protein_id="AAH08727.1"
/db_xref="GI:14250550"
/translation="MAGNSSSCPTPGGHLNGYVPVYAFFPMLGLSPGALTTLL
QHLPVSGYSTPATIETQSSSEIEVPSPPPLPRIVKPCVCDKSSGYHGV
SACBCKGFFRRSRKOKNVTCHRDKNKIIINKVNRKOYCRLOKCFVGMKSEYRN
DRNKKKEVPKPESESTLTPVEGELIEKVRKHAHOETFPALCQKGYTNNSQORV
SLDLDLWDFSELTKCIITKTFEPAKQPLFTLTIAQDITLLKAACDLILIRICTR
YTPQDFTMTFSDGLTNTARTOMHNAFGPLTDLVFAFANQLPLEMDAETGLLSAICL
ICGDRDLEQPDVDRVMDLQEPLEALKVYVRKRRSRPHMFPMKMKITDRLRISAKGA
ERVITLAKMEIPGSMPLPIQIEMLENSGLDLSGPGGGGRLGGLAPPGPSPLSLP
SSNRPSPATHSP"
BASE COUNT 510 a 822 c 672 g 437 t
ORIGIN
Query Match 85.5%; Score 1783.2; DB 9; Length 2441;
Best Local Similarity 99.8%; Pred. No. 4.3e-313;
Matches 1796; Conservative 0; Mismatches 3; Indels 1; Gaps 1;
QY 288 CTCCATTGAGACCCAGACGAGCAGTCTGAAGATAGTCCCGAGCCCTCCCTCGCCACC 347
|
|
|
Db 641 CACCATTGAGACCCAGACGAGCAGTCTGAAGATAGTCCCGAGCCCTCCCTCGCCACC 700
|
|
|
QY 348 CCTCTACCCCGCATCTACAGCCTTGTCTGTCTGTCAGGACAAGTCTCAGGCTACCA 407
|
|
|
Db 701 CCTCTACCCCGCATCTACAGCCTTGTCTGTCTGTCAGGACAAGTCTCAGGCTACCA 760
|
|
|
QY 408 CTATGGGCTGACGCGCTGTGAGGCTGCAAGGGCTTCTTCCCGCGAGCATCCAGAGAA 467
|
|
|
Db 761 CTATGGGCTGACGCGCTGTGAGGCTGCAAGGGCTTCTTCCCGCGAGCATCCAGAGAA 820
|
|
|
QY 468 CATGTGTACACGCTGTACCCGGGACAAAGTGCATCATCAACAAGTGCACCCGGAACCC 527
|
|
|
Db 821 CATGTGTACACGCTGTACCCGGGACAAAGTGCATCATCAACAAGTGCACCCGGAACCC 880
|
|
|
QY 528 CTGCCAGTACTGCCGACGTGCAGAAAGTGTCTTGAAGTGGGATGTCTCAAGAGTCTGTAG 587
|
|
|
Db 881 CTGCCAGTACTGCCGACGTGCAGAAAGTGTCTTGAAGTGGGATGTCTCAAGAGTCTGTAG 940
|
|
|
QY 588 AAACGACCGGAACAGAGAGAGAGAGAGTGCCTCCAGCCGAGTCTCTGAGAGCTACAC 647
|
|
|
Db 941 AAACGACCGGAACAGAGAGAGAGAGTGCCTCCAGCCGAGTCTCTGAGAGCTACAC 1000
|
|
|
QY 648 GCTGACGCGGAGGTGGGGAGTCTATTGAGAAGTGTGCGAAAGCGCACAGGAAACCTT 707
|
|
|
Db 1001 GCTGACGCGGAGGTGGGGAGTCTATTGAGAAGTGTGCGAAAGCGCACAGGAAACCTT 1060
|
|
|
QY 708 CCTGCGCTCTGCCAGTGGGCAATATACACTACGAACAACAGCTCAGACAACAGTGTCTC 767
|
|
|
Db 1061 CCTGCGCTCTGCCAGTGGGCAATATACACTACGAACAACAGCTCAGACAACAGTGTCTC 1120
|
|
|
QY 768 TCTGGACATTGACCTCTGGGACAAGTTCAGTGAACCTCTCCACCAAGTGCATTAAGAC 827
|
|
|
Db 1121 TCTGGACATTGACCTCTGGGACAAGTTCAGTGAACCTCTCCACCAAGTGCATTAAGAC 1180
|
|
|
QY 828 TGTGAGTTCGCAAGAGCTGCCCGGCTTACCACCTTCCATCCGCGAGCATCATC 887
|
|
|
Db 1181 TGTGAGTTCGCAAGAGCTGCCCGGCTTACCACCTTCCATCCGCGAGCATCATC 1240
|
|
|
QY 888 CCTCTCAAGGCTGCTGCCGTGACATCCTGTCTCTCGGATCTGCACCGGCTACAGCC 947
|
|
|
Db 1241 CCTCTCAAGGCTGCTGCCGTGACATCCTGTCTCTCGGATCTGCACCGGCTACAGCC 1300
|
|
|
QY 948 CGAGCAGGACCATGACCTTCTCGGACGGGCTGACCTTAACCGGACCCAGATGACAA 1007
|
|
|
Db 1301 CGAGCAGGACCATGACCTTCTCGGACGGGCTGACCTTAACCGGACCCAGATGACAA 1360
|
|
|
QY 1008 CGTGGCTTCGGCCCGCTCAGGACCTGTCTTTCCTTCGCGCAACAGCTGTCTCCCT 1067
|
|
|
```

Db 1361 CGCTGGCTTGGCCCTCCACGACCTGGTCTTTGCCCTTCCGCAACCAAGCTCTGCTCCCT 1420  
Qy 1068 GGAGATGATGATCGGAGACGGGCTGCTCAGCGCCATCTCCTCCTCATCTCGGAGACGG 1127  
Db 1421 GGAGATGATGATCGGAGACGGGCTGCTCAGCGCCATCTCCTCCTCATCTCGGAGACGG 1480  
Qy 1128 CCAGACCTGGAGACGCGGACCGGGTGGACATGCTCAGGAGCGGCTGCTCGAGGCGCT 1187  
Db 1481 CCAGACCTGGAGACGCGGACCGGGTGGACATGCTCAGGAGCGGCTGCTCGAGGCGCT 1540  
Qy 1188 AAGGCTTACCTGGAGACGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1247  
Db 1541 AAGGCTTACCTGGAGACGCGGACCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1600  
Qy 1248 GAAATTAATCTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGAGCGGGTGTATCACCTGAA 1307  
Db 1601 GAAGTAATCTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGAGCGGGTGTATCACCTGAA 1660  
Qy 1308 GATGAGATCCCGGCTCCATGCCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGG 1367  
Db 1661 GATGAGATCCCGGCTCCATGCCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGGG 1720  
Qy 1368 CTTGACACTCTGAGCGACAGCGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1427  
Db 1721 CTTGACACTCTGAGCGACAGCGGGGGTGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG 1780  
Qy 1428 CCGCCAGGACGCTGAGCGCCAGGCTCAGCGCCAGCTCCACAGAGAGCGGCGGCGAC 1487  
Db 1781 CCGCCAGGACGCTGAGCGCCAGGCTCAGCGCCAGCTCCACAGAGAGCGGCGGCGAC 1840  
Qy 1488 CCACTCCCGGTGACCGCCAGGCTGACAGCGGCTCAGCGCCAGCTCCACAGAGAGCGG 1547  
Db 1841 CCACTCCCGGTGACCGCCAGGCTGACAGCGGCTCAGCGCCAGCTCCACAGAGAGCGG 1900  
Qy 1548 CTCTGCTTTTACGAGACCTGTCACCGGACGAGCGGCTGCGCCCTCAGCGGCTTTC 1607  
Db 1901 CTCTGCTTTTACGAGACCTGTCACCGGACGAGCGGCTGCGCCCTCAGCGGCTTTC 1960  
Qy 1608 GCAGTACTGGGACCTTCCCTGGGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1667  
Db 1961 GCAGTACTGGGACCTTCCCTGGGGAGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2020  
Qy 1668 AGGCTTGGGCTCAGTGGAGCTGCTGCTCCACAGCGCTGGGCTGAGCTCAGAGGCGGAG 1727  
Db 2021 AGGCTTGGGCTCAGTGGAGCTGCTGCTCCACAGCGCTGGGCTGAGCTCAGAGGCGGAG 2080  
Qy 1728 GCGAGAACTGAGTGGGCTTCTGCTGGGTCTCAGATGGGTCTCGGGGGCTCGTG 1787  
Db 2081 GCGAGAACTGAGTGGGCTTCTGCTGGGTCTCAGGATGGGTCTCGGGGGCTCGTG 2140  
Qy 1788 TTCATCAGACACCCCTCTGCGGAGCTCAGCAGCTCCTCATCAGCAGCAAGCGGCGAG 1847  
Db 2141 TTCATCAGACACCCCTCTGCGGAGCTCAGCAGCTCCTCATCAGCAGCAAGCGGCGAG 2200  
Qy 1848 TTGGCTCCCGCATCTCAGAACTCACAAGCCTGCTCCCGAGCTGGGAGCCTCAACCT 1907  
Db 2201 TTGGCTCCCGCATCTCAGAACTCACAAGCCTGCTCCCGAGCTGGGAGCCTCAACCT 2260  
Qy 1908 CCGCCCTGCTGCTGGTGGACAGAGGGGGTGGAGAGGGGGGGGGGGGGGGGGGGGGGG 1967  
Db 2261 CCGCCCTGCTGCTGGTGGACAGAGGGGGTGGAGAGGGGGGGGGGGGGGGGGGGGGGG 2320  
Qy 1968 ATACCCCTGCCATACCAACCCAGGTATTAATCTCGCTGGTGTGTTTTTAAATTTTAA 2026  
Db 2321 ATACCCCTGCCATACCAACCCAGGTATTAATCTCGCTGGTGTGTTTTTAAATTTTAA 2380  
Qy 2027 TTTTGTGTTTTTATTAAGAAATTTTCAATTTTAAAGCACAATAAAAAAAAAAAAAA 2086  
Db 2381 TTTTGTGTTTTTATTAAGAAATTTTCAATTTTAAAGCACAATAAAAAAAAAAAAAA 2440

RESULT 3  
AR061121

LOCUS: AR061121 3036 bp DNA linear PAT 29-SEP-1999  
DEFINITION: Sequence 1 from patent US 5843642.  
ACCESSION: AR061121  
VERSION: AR061121.1 GI:5988812  
KEYWORDS: Unknown.  
SOURCE: Unknown.  
ORGANISM: Unclassified.  
REFERENCE: 1 (bases 1 to 3036)  
AUTHORS: Dmitrovsky, E., Warrell, R. P. Jr., Miller, W. H. Jr. and Frankel, S.  
TITLE: Methods for detection of acute promyelocytic leukemia (APL)  
JOURNAL: Patent: US 5843642-A 1 01-DEC-1998;  
FEATURES: Location/Qualifiers  
1..3036  
BASE COUNT: 605 a 1018 c 890 g 523 t  
ORIGIN: /organism="unknown"  
Query Match 85.4%; Score 1782.2; DB 6; Length 3036;  
Best Local Similarity 99.8%; Pred. No. 7.2e-313;  
Matches 1784; Conservative 0; Mismatches 3; Indels 0; Gaps 0;  
Qy 290 CCATTGAGACCCAGAGCAGCAGTTCTGAAGAGATAGTGCACGCTCTCCGCGCACCC 349  
Db 1250 CCATTGAGACCCAGAGCAGCAGTTCTGAAGAGATAGTGCACGCTCTCCGCGCACCC 1309  
Qy 350 CTCCTACCCGGCATCTACAAGCCTTGGTTGTCTGTGAGACAGTCTCTAGGCTACCACT 409  
Db 1310 CTCCTACCCGGCATCTACAAGCCTTGGTTGTCTGTGAGACAGTCTCTAGGCTACCACT 1369  
Qy 410 ATGGGTGTCAGCGCTGTGAGGGCTGCAAGGGCTTCTTCGCGCCAGCATCCAGAAGAAC 469  
Db 1370 ATGGGTGTCAGCGCTGTGAGGGCTGCAAGGGCTTCTTCGCGCCAGCATCCAGAAGAAC 1429  
Qy 470 TGGTGTACAGCTGTCAACGGGACAAAGACTGCATCATCAACAGGTGACCGGAACCCCT 529  
Db 1430 TGGTGTACAGCTGTCAACGGGACAAAGACTGCATCATCAACAGGTGACCGGAACCCCT 1489  
Qy 530 GCAGTACTGCGGCTGTCAGAGAGTCTTGAAGTGGGCTGTCCAAAGGAGTCTGTGAGAA 589  
Db 1490 GCAGTACTGCGGCTGTCAGAGAGTCTTGAAGTGGGCTGTCCAAAGGAGTCTGTGAGAA 1549  
Qy 590 AGCAGCGAAACAAAGAGAGAGGAGTGCCTCAAGCCGAGTCTCTGAGAGCTACACGC 649  
Db 1550 AGCAGCGAAACAAAGAGAGAGGAGTGCCTCAAGCCGAGTCTCTGAGAGCTACACGC 1609  
Qy 650 TGACCGCGGAGTGGGGAGCTCATTTGAGAGGTGCGCAAGCGCACAGGAAACCTTCC 709  
Db 1610 TGACCGCGGAGTGGGGAGCTCATTTGAGAGGTGCGCAAGCGCACAGGAAACCTTCC 1669  
Qy 710 CTGCGCTCTGCGGCTGGGCAATACACTACGACAAACAGCTCAGACACAGCTGTCTCTC 769  
Db 1670 CTGCGCTCTGCGGCTGGGCAATACACTACGACAAACAGCTCAGACACAGCTGTCTCTC 1729  
Qy 770 TGGACATTCACCTCTGGGCAAGTTTCAGTGAATCTCTCCACCAAGTGCATCATTAAGACTG 829  
Db 1730 TGGACATTCACCTCTGGGCAAGTTTCAGTGAATCTCTCCACCAAGTGCATCATTAAGACTG 1789  
Qy 830 TGGAGTTGCGCAAGCAGCTGCCCGCTTCCACCACTTCCACCGGCGGAGTACAGTACCC 889  
Db 1790 TGGAGTTGCGCAAGCAGCTGCCCGCTTCCACCACTTCCACCGGCGGAGTACAGTACCC 1849  
Qy 890 TCCTCAAGGCTGCTGCTGGGACATCTCTGCGGATCTGCGGCTGACGCGGTACACGCGCG 949  
Db 1850 TCCTCAAGGCTGCTGCTGGGACATCTCTGCGGATCTGCGGCTGACGCGGTACACGCGCG 1909  
Qy 950 AGCAGGACACCATGACCTTCTCGGAGGGCTGACCTCTGAACCGGAGCCAGATGCAACG 1009  
Db 1910 AGCAGGACACCATGACCTTCTCGGAGGGCTGACCTCTGAACCGGAGCCAGATGCAACG 1969  
Qy 1010 CTGGCTTGGGCGGCTTCCACCGACCTGGTGTGCTTTCGCTTCCGCAACAGCTGCTGCCCTGG 1069  
Db 1970 CTGGCTTGGGCGGCTTCCACCGACCTGGTGTGCTTTCGCTTCCGCAACAGCTGCTGCCCTGG 2029



QY	1070	AGATGGATGATGGAGACGGGGCTCTCAGCGCCATCTGCCTCATCTCGGAGACGGCC	1129
Db	2030	AGATGGATGATGGAGACGGGGCTCTCAGCGCCATCTGCCTCATCTCGGAGACGGCC	2089
QY	1130	AGGACCTGGAGACGGGACCGGGTGGATGCTGAGGAGCGCTGCTGGAGCGCTAA	1189
Db	2090	AGGACCTGGAGACCGGACCGGGTGGATGCTGAGGAGCGCTGCTGGAGCGCTAA	2149
QY	1190	AGGTCTACGTGGGAAGCGGAGCGCCAGCGCCGCCACATGTTCCCAAGATGCTAATGA	1249
Db	2150	AGGTCTACGTGGGAAGCGGAGCGCCAGCGCCGCCACATGTTCCCAAGATGCTAATGA	2209
QY	1250	AGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGAGCGGGTGCATCACCTGNAGA	1309
Db	2210	AGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGCTGAGCGGGTGCATCACCTGNAGA	2269
QY	1310	TGAGATCCCGGGCTCCATGCCGCTCTCTATCCAGAAATGTTGGAGAACTCAGAGGGCC	1369
Db	2270	TGAGATCCCGGGCTCCATGCCGCTCTCTATCCAGAAATGTTGGAGAACTCAGAGGGCC	2329
QY	1370	TGGACACTCTGACGGACAGCGGGGGTGGGGGGGGGAGCGGGGTGGCTGCCCGCCC	1429
Db	2330	TGGACACTCTGACGGACAGCGGGGGTGGGGGGGGGAGCGGGGTGGCTGCCCGCCC	2389
QY	1430	CGCAGGACAGCTGTAGCCCGCAGCTCAGCCCGCAGCTCCAAAGAGAGCGCGGCCACCC	1489
Db	2390	CGCAGGACAGCTGTAGCCCGCAGCTCAGCCCGCAGCTCCAAAGAGAGCGCGGCCACCC	2449
QY	1490	ACTCCCGTGACCGCCACGCCACATGGACAGCGCTCGCCCTCGCCCGCGCTTTCT	1549
Db	2450	ACTCCCGTGACCGCCACGCCACATGGACAGCGCTCGCCCTCGCCCGCGCTTTCT	2509
QY	1550	CTGCCCTTTCTACCGACATGTGACCCCGCACAGCCCTGCCCGCACCTGCCCTCCCGGGC	1609
Db	2510	CTGCCCTTTCTACCGACATGTGACCCCGCACAGCCCTGCCCGCACCTGCCCTCCCGGGC	2569
QY	1610	AGTACTGGGACCTTCCCTGGGGAGCGGGGAGGAGGAGGAGCGACTCTTGACAGAG	1669
Db	2570	AGTACTGGGACCTTCCCTGGGGAGCGGGGAGGAGGAGGAGGAGCGACTCTTGACAGAG	2629
QY	1670	GCTGGGCGCTCAGTGGACTGCTGCTGCCACAGCGCTGGGCTCAGCTCAGGCGCGAGGC	1729
Db	2630	GCTGGGCGCTCAGTGGACTGCTGCTGCCACAGCGCTGGGCTCAGCTCAGGCGCGAGGC	2689
QY	1730	CAGGAATGAGTGGGCGCTGCTGCTGCTCAGGATGGGCTCCTGGGGCGCTCGTGT	1789
Db	2690	CAGGAATGAGTGGGCGCTGCTGCTGCTCAGGATGGGCTCCTGGGGCGCTCGTGT	2749
QY	1790	CATCAAGACACCCCTTGCCCGAGCTCACCACATCTTTCATCACCAGCAAAAGCGGACTT	1849
Db	2750	CATCAAGACACCCCTTGCCCGAGCTCACCACATCTTTCATCACCAGCAAAAGCGGACTT	2809
QY	1850	GGCTCCCGCATCTCAGAACTCAGACCATGCTCCCGAGCTGGGAACTCAACCTCC	1909
Db	2810	GGCTCCCGCATCTCAGAACTCAGACCATGCTCCCGAGCTGGGAACTCAACCTCC	2869
QY	1910	CCCTGCCCTGGTGTGACAGAGGGGTGGGACAGGGGGGGGGTTCCTCCCTGTACAT	1969
Db	2870	CCCTGCCCTGGTGTGACAGAGGGGTGGGACAGGGGGGGGGTTCCTCCCTGTACAT	2929
QY	1970	ACCTGCCATACCAACCCAGGATTAATTTCTCGCTGTTTGTATTTTAAATTTT	2029
Db	2930	ACCTGCCATACCAACCCAGGATTAATTTCTCGCTGTTTGTATTTTAAATTTT	2989
QY	2030	TTGTTTTGATTTTTTAATAAGAAATTTTCAATTTTAAGCAAAAAA	2076
Db	2990	TTGTTTTGATTTTTTAATAAGAAATTTTCAATTTTAAGCAAAAAA	3036
RESULT 4	196215		
LOCUS		3036 bp	DNA linear PAT 01-DEC-1998

DEFINITION	Sequence 52 from patent US 5734039.		
ACCESSION	196215		
VERSION	196215.1 GI:3940685		
KEYWORDS	Unknown.		
SOURCE	Unknown.		
ORGANISM	Unclassified.		
REFERENCE	1 (bases 1 to 3036)		
AUTHORS	Calabretta,B. and Skorski,T.		
TITLE	Antisense oligonucleotides targeting cooperating oncogenes		
JOURNAL	Patent: US 5734039-A 52 31-MAR-1998;		
FEATURES	Location/Qualifiers		
source	1..3036		
BASE COUNT	605 a	1018 c	890 g 523 t
ORIGIN	/organism="unknown"		
Query Match	85.4%; Score 1782.2; DB 6; Length 3036;		
Best Local Similarity	99.8%; Pred. No. 7.2e-313;		
Matches 1784; Conservative	0;	Mismatches 3;	Indels 0; Gaps 0;
QY	290 CCATTGAGACCCAGAGCAGCAGTCTCTGAAGAGATAGTCCCGCAGCCCTCCTCGCCACCCC	349	
Db	1250 CCATTGAGACCCAGAGCAGCAGTCTCTGAAGAGATAGTCCCGCAGCCCTCCTCGCCACCCC	1309	
QY	350 CTCTACCCCGCATCTACAAGCCTTGTCTGTCTGAGGACAAAGTCTCAGCTACCACT	409	
Db	1310 CTCTACCCCGCATCTACAAGCCTTGTCTGTCTGAGGACAAAGTCTCAGCTACCACT	1369	
QY	410 ATGGGGTCAGCGCCTGTGAGGGCTGCAAGGGCTTCTTCCGCGCAGCATCCAGAAAGACA	469	
Db	1370 ATGGGGTCAGCGCCTGTGAGGGCTGCAAGGGCTTCTTCCGCGCAGCATCCAGAAAGACA	1429	
QY	470 TGTGTACAGTGTACCGGGACAAAGAACTGCATCATCAACAAGGTGACCCGGAACCCCT	529	
Db	1430 TGTGTACAGTGTACCGGGACAAAGAACTGCATCATCAACAAGGTGACCCGGAACCCCT	1489	
QY	530 GCCAGTACTGCCGACTCGAGAAGTGTCTTGAAGTGGGATGTCCCAAGGAGTCTGTAGAA	589	
Db	1490 GCCAGTACTGCCGACTCGAGAAGTGTCTTGAAGTGGGATGTCCCAAGGAGTCTGTAGAA	1549	
QY	590 ACGACCGAAACAAAGAAAGAGAGGTGCGCAAGCGGAGTGTCTGAGAGCTACACGC	649	
Db	1550 ACGACCGAAACAAAGAAAGAGAGGTGCGCAAGCGGAGTGTCTGAGAGCTACACGC	1609	
QY	650 TGACCGCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGCAAGCGCAGCAGGAAACCTTCC	709	
Db	1610 TGACCGCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGCAAGCGCAGCAGGAAACCTTCC	1669	
QY	710 CTGCGCTCTGCGCAGCTGGGCAAAATACACTAGCAACAGCTCAGAACAACTGTCTCTC	769	
Db	1670 CTGCGCTCTGCGCAGCTGGGCAAAATACACTAGCAACAGCTCAGAACAACTGTCTCTC	1729	
QY	770 TGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTTCCACCAAGTGCATCATTAAGACTG	829	
Db	1730 TGGACATTGACCTCTGGGACAAAGTTCAGTGAACCTTCCACCAAGTGCATCATTAAGACTG	1789	
QY	830 TGGAGTTGCGCAAGCAGCTGCCGGCTTCAACACCTCACCCTCCCGACCAAGATCACCC	889	
Db	1790 TGGAGTTGCGCAAGCAGCTGCCGGCTTCAACACCTCACCCTCCCGACCAAGATCACCC	1849	
QY	890 TCCTCAAGGCTGCTGCTGGACATCTCTGATCTCTGCGGATCTGCAACGGGTACACGGCG	949	
Db	1850 TCCTCAAGGCTGCTGCTGGACATCTCTGATCTCTGCGGATCTGCAACGGGTACACGGCG	1909	
QY	950 AGCAGGACACCATGACCTTCTCGGACGGCTGACCTCAACCGGACCCAGATGCACACG	1009	
Db	1910 AGCAGGACACCATGACCTTCTCGGACGGCTGACCTCAACCGGACCCAGATGCACACG	1969	
QY	1010 CTGGCTTCGGCCCTCACCAGCTGCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC	1069	
Db	1970 CTGGCTTCGGCCCTCACCAGCTGCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTC	2029	

QY 1070 AGATGGATGATCCGAGACGGGGTGTCTAGCGCCATCTGCTCATCTGCGGAGACCGCC 1129  
 Db 2030 AGATGGATGATCGGAGACGGGGTGTCTAGCGCCATCTGCTCATCTGCGGAGACCGCC 2089  
 QY 1130 AGGACCTGGAGACGGCGGATGACATGCTGACGAGCCGCTGCTGAGGCGCTAA 1189  
 Db 2090 AGGACCTGGAGACGGCGGATGACATGCTGACGAGCCGCTGCTGAGGCGCTAA 2149  
 QY 1190 AGGTCTAGCTCGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAATGA 1249  
 Db 2150 AGGTCTAGCTCGGAAGCGGAGCGCCAGCGCCCGCCACATGTTCCCAAGATGCTAAATGA 2209  
 QY 1250 AGATTACTGACTCGGAAGCATAGCGCCAAAGGGGGTGAAGGGGTGATCAGCTGAAGA 1309  
 Db 2210 AGATTACTGACTCGGAAGCATAGCGCCAAAGGGGGTGAAGGGGTGATCAGCTGAAGA 2269  
 QY 1310 TGGAGATCCCGGGTCCATGCGCGCTCTCATCCAGGAATGTTGGAGACTCAGAGGCC 1369  
 Db 2270 TGGAGATCCCGGGTCCATGCGCGCTCTCATCCAGGAATGTTGGAGACTCAGAGGCC 2329  
 QY 1370 TGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGGAGCGGGTGGCTGCCCGCCC 1429  
 Db 2330 TGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGGAGCGGGTGGCTGCCCGCCC 2389  
 QY 1430 CGCCAGGACGCTGTAGCCCGCAGCTCAGCCCGCAGCTCCACAGAGCAGCGCGCCACCC 1489  
 Db 2390 CGCCAGGACGCTGTAGCCCGCAGCTCAGCCCGCAGCTCCACAGAGCAGCGCGCCACCC 2449  
 QY 1490 ACTCCCGTGAACCGCCAGCGCCGAGCGGGGGTGGGGGCGGGAGCGGGTGGCTGCCCGCCC 1549  
 Db 2450 ACTCCCGTGAACCGCCAGCGCCGAGCGGGGGTGGGGGCGGGAGCGGGTGGCTGCCCGCCC 2509  
 QY 1550 CTGCGTTTCTACCGACCATGTGACCCCGCAGCGCCCTGCCCGCCAGCTGCCCTCCCGGGC 1609  
 Db 2510 CTGCGTTTCTACCGACCATGTGACCCCGCAGCGCCCTGCCCGCCAGCTGCCCTCCCGGGC 2569  
 QY 1610 AGTACTGGGACCTTCCCTGGGGACCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1669  
 Db 2570 AGTACTGGGACCTTCCCTGGGGACCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 2629  
 QY 1670 GCTTGGGCGCTCAGTGACGCTGCTGCTCCACAGCCTGGGCTGAGCTCAGAGGCGGAGGC 1729  
 Db 2630 GCTTGGGCGCTCAGTGACGCTGCTGCTCCACAGCCTGGGCTGAGCTCAGAGGCGGAGGC 2689  
 QY 1730 CAGGAATGAGTGAGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1789  
 Db 2690 CAGGAATGAGTGAGGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2749  
 QY 1790 CATCAGACACCCCTGCGCCAGCTCAGCAGCTCAGCAGCTCAGCAGCTCAGCAGCTCAGCAGCT 1849  
 Db 2750 CATCAGACACCCCTGCGCCAGCTCAGCAGCTCAGCAGCTCAGCAGCTCAGCAGCTCAGCAGCT 2809  
 QY 1850 GGCTCCCGCATCTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTC 1909  
 Db 2810 GGCTCCCGCATCTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTCAGAACTC 2869  
 QY 1910 CCCTGCTCGGTGGTGACAGAGGGGGTGGGACAGGCGGGGGTGGGCTGACAT 1969  
 Db 2870 CCCTGCTCGGTGGTGACAGAGGGGGTGGGACAGGCGGGGGTGGGCTGACAT 2929  
 QY 1970 ACCCTGCCATACCAACCCAGGATTAATTTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2029  
 Db 2930 ACCCTGCCATACCAACCCAGGATTAATTTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2989  
 QY 2030 TTGTTTTGATTTTAAATAAGATTTTCAATTTAAGCAAAAAA 2076  
 Db 2990 TTGTTTTGATTTTAAATAAGATTTTCAATTTAAGCAAAAAA 3036

## RESULT 5

HUMPMRLAR

LOCUS

DEFINITION

Human PML-RAR protein (PML-RAR) mRNA, complete cds.  
 3036 bp mRNA linear PRI 08-JAN-1995

## ACCESSION

M73779.1 GI:190124

PML-RAR protein; fusion protein.

SOURCE Homo sapiens cDNA to mRNA.

ORGANISM Homo sapiens

REFERENCE  
 1 (bases 1 to 3036)  
 Kakizuka, A., Miller, W.H. Jr., Umesono, K., Warrell, R.P. Jr.,  
 Frankel, S.R., Murty, V.V., Dmitrovsky, E. and Evans, R.M.

TITLE  
 Chromosomal translocation t(15;17) in human acute promyelocytic  
 leukemia fuses RAR alpha with a novel putative transcription  
 factor, PML

JOURNAL  
 Cell 66 (4), 663-674 (1991)

MEDLINE  
 91347368

PUBMED  
 1652368

## FEATURES

source

Location/Qualifiers

1..3036

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/map="chromosome t(15;17)"

/cell\_type="bone marrow"

1..3036

/gene="PML-RAR"

67..2460

/gene="PML-RAR"

/codon\_start=1

/product="PML-RAR protein"

/protein\_id="AAA60126.1"

/db\_xref="GI:190125"

/translation="MEPAPARSPRPODPAPOPTMPPTTSPGQSPSPSPPTER

APASEFEFLRCQCOEAKPKLLCLHTLCSGLEASGMCQPCICQAPNPLADTP

ALNVFESLQRLSVYQIVDAQVTRCKESADFCFECEQLLCAKCFPAHOMFLK

HEARPLAELRNQSVREFLDGRTKNTFNTPSNHRTPLTISIYCRGSKPKCCSCALL

DSHSELKCDISAEIQORQEBLDMATQALQDSAFGAHQAHAQVQLGARAETE

ELJRVVQVAVHVRQERELLEAVDARYQDYEMASRLGRDLVQLRIQTGSAIVO

RMKYASDOEVLDMHGLFRLQALFLROEPQSLQAAVDTGDFDKFLQLDSSCITO

KALETQSSSEELVPSPPPLPRIYKFCVQDKASGSHYGVACGCKGKFFRRS

IQKNMYTCHRNKNIIVKTRNRCQYCLQKCFEVMGSKESVRNDRNKKKVEPKPE

CSBSYTLTPEVGLIEIKVRAHOETFPALCOLGYTNNSEQRVSLDIDLWDFSEL

STKIIKTVEFALQFGFTLTADQITLLKACGLDILIRICTRYTPEQDTMTFSDG

LTNRTQMHNAGFPLTDLVAFANOLLPLEMDDAETGLLSAICLCGDRDLQPDPR

VDMLQPLLEALAVYVRKRPSPHMPKMLKTTDLRSTISAKGAERVTLKMEIPGS

MPPLQEMLENSGLDTLSGPGGGGRDGLAPPGSCPSLSPSSNRSPATHSP"

BASE COUNT 505 a 1018 c 890 g 523 t

ORIGIN chromosome t(15;17) translocation.

## Query Match

Best Local Similarity 85.4%; Score 1782.2; DB 9; Length 3036;

Mismatches 1784; Conservative 0; Indels 3; Gaps 0;

QY 290 CCATTGAGACCCAGAGCAGCAGTTCCTGAAGAGATAGTGCCAGCCCTCCCTCGCACCCC 349

Db 1250 CCATTGAGACCCAGAGCAGCAGTTCCTGAAGAGATAGTGCCAGCCCTCCCTCGCACCCC 1309

QY 350 CTCTACCCCGCATCTACAAGCCTTGTCTCTCAGGACAACTCTCAGGCTACCACT 409

Db 1310 CTCTACCCCGCATCTACAAGCCTTGTCTCTCAGGACAACTCTCAGGCTACCACT 1369

QY 410 ATGGGCTCAGCGCTGTGAGGGCTGCAAGGGCTTCTTCGCGCGCAGCATCCAGAGACA 469

Db 1370 ATGGGCTCAGCGCTGTGAGGGCTGCAAGGGCTTCTTCGCGCGCAGCATCCAGAGACA 1429

QY 470 TGGTGTACAGTGTCAACCGGACAAAGTGCATCAACAAGTGCACCGGACCCCT 529

Db 1430 TGGTGTACAGTGTCAACCGGACAAAGTGCATCAACAAGTGCACCGGACCCCT 1489

QY 530 GCCAGTACTGCCGACTGCGAAGTGTCTTGAAGTGGGCATGTCCAAGAGTCTGTGAGAA 589

Db 1490 GCCAGTACTGCCGACTGCGAAGTGTCTTGAAGTGGGCATGTCCAAGAGTCTGTGAGAA 1549

QY 590 ACCAGCGAAACAGAGAGAGAGTGCCCAAGCCCGAGTGTCTGAGAGCTACAGC 649

|||||

Db 1550 ACGACCAACAAAGAGAGAGAGGTGCCCAAGCCGAGTCTCTCAGAGCTACACGC 1609  
Qy 650 TGAGCCGCGAGGTGGGAGAGCTCATTCAGAAAGTTCGCAAGCGCACAGGAAACCTTCC 709  
Db 1610 TGAGCCGCGAGGTGGGAGAGCTCATTCAGAAAGTTCGCAAGCGCACAGGAAACCTTCC 1669  
Qy 710 CTGCCCCCTGCGAGCTGGGAAATACACTAGCAACAGCTCAGAACAGGTGCTCTC 769  
Db 1670 CTGCCCCCTGCGAGCTGGGAAATACACTAGCAACAGCTCAGAACAGGTGCTCTC 1729  
Qy 770 TGGACATTGACCTCTGGGACAAAGTTCACTGAACTCTCCACCAAGTGCATCAAGAGCTG 829  
Db 1730 TGGACATTGACCTCTGGGACAAAGTTCACTGAACTCTCCACCAAGTGCATCAAGAGCTG 1789  
Qy 830 TGGAGTTTCGCAACAGCTGCGCGCTTCACCAACCTCACCATCGCCGACCAAGATCAACC 889  
Db 1790 TGGAGTTTCGCAACAGCTGCGCGCTTCACCAACCTCACCATCGCCGACCAAGATCAACC 1849  
Qy 890 TCCTCAAGGTGCTGCTGGACATCTCTGATCTCTGCGGATCTGCAAGCGGTACAGCCGC 949  
Db 1850 TCCTCAAGGTGCTGCTGGACATCTCTGATCTCTGCGGATCTGCAAGCGGTACAGCCGC 1909  
Qy 950 AGCAGGACCATGACCTCTCGGACGCGCTGACCTGAACCGGACCCAGATGCAACAG 1009  
Db 1910 AGCAGGACCATGACCTCTCGGACGCGCTGACCTGAACCGGACCCAGATGCAACAG 1969  
Qy 1010 CTGCTTTGCGCCCTCACCACCTGCTTTGCTCTCGCCCAACAGCTGCTGCCCTGG 1069  
Db 1970 CTGCTTTGCGCCCTCACCACCTGCTTTGCTCTCGCCCAACAGCTGCTGCCCTGG 2029  
Qy 1070 AGATGGATGATCGGAGACGCGGCTGCTCAGCGCATCTGCTCATCTCGGAGACCGCC 1129  
Db 2030 AGATGGATGATCGGAGACGCGGCTGCTCAGCGCATCTGCTCATCTCGGAGACCGCC 2089  
Qy 1130 AGGACCTGGAGACGCGGCTGGACATCTGCTCAGGACCGCTGCTGGAGGCGCTAA 1189  
Db 2090 AGGACCTGGAGACGCGGCTGGACATCTGCTCAGGACCGCTGCTGGAGGCGCTAA 2149  
Qy 1190 AGGTCTAGTCTCGGAAGCGGAGCGCCAGCGCCGCCACATGTTCGCCAAGATGCTAAATGA 1249  
Db 2150 AGGTCTAGTCTCGGAAGCGGAGCGCCAGCGCCGCCACATGTTCGCCAAGATGCTAAATGA 2209  
Qy 1250 AGATTACTGACCTCGGAAGCATACGCGCCAAAGGGGCTGAGCGGTGATCAGCTGAAGA 1309  
Db 2210 AGATTACTGACCTCGGAAGCATACGCGCCAAAGGGGCTGAGCGGTGATCAGCTGAAGA 2269  
Qy 1310 TGGAGATCCCGGCTCCATGCGGCTCTCATCCAGGAATGTGGAGAACTCAGAGGGCC 1369  
Db 2270 TGGAGATCCCGGCTCCATGCGGCTCTCATCCAGGAATGTGGAGAACTCAGAGGGCC 2329  
Qy 1370 TGGACACTCTGAGCGGACAGCGGGGCTGGGGGCGGAGCGGCTGCGCTGCCCGCC 1429  
Db 2330 TGGACACTCTGAGCGGACAGCGGGGCTGGGGGCGGAGCGGCTGCGCTGCCCGCC 2389  
Qy 1430 CGCCAGGAGCTGTAGCCCCAGCCTCAGCCCCAGCTTCCAAAGAGAGCAGCCCGGCCACC 1489  
Db 2390 CGCCAGGAGCTGTAGCCCCAGCCTCAGCCCCAGCTTCCAAAGAGAGCAGCCCGGCCACC 2449  
Qy 1490 ACTCCCGGTGACCGCCAGCCACATGAGACAGCGCTCGCCCTCGCGCCCGCGCTTTTCT 1549  
Db 2450 ACTCCCGGTGACCGCCAGCCACATGAGACAGCGCTCGCCCTCGCGCCCGCGCTTTTCT 2509  
Qy 1550 CTGCTTTTCTACGACCATGTACCCCGCCAGCCCTGCCCGCCCTGCCCTGCCCGGC 1609  
Db 2510 CTGCTTTTCTACGACCATGTACCCCGCCAGCCCTGCCCGCCCTGCCCTGCCCGGC 2569  
Qy 1610 AGTACTGGGAGCTTCTCTGGGGAGCGGAGGAGGAGCGAGCTCTCTTGGACAGAG 1669  
Db 2570 AGTACTGGGAGCTTCTCTGGGGAGCGGAGGAGGAGGAGCGAGCTCTCTTGGACAGAG 2629  
Qy 1670 GCCTGGGCGCTCAGTGGAGTGGCTGTCTCCACAGCGCTGGGCTGACGTACAGAGCCGAGCC 1729  
Db 2630 GCCTGGGCGCTCAGTGGAGTGGCTGTCTCCACAGCGCTGGGCTGACGTACAGAGCCGAGCC 2689

Qy 1730 CAGGAACCTGAGTGAGGCCCTCTGCTGCTCAGGATGGTCTCGGGGCGCTCGTGT 1789  
Db 2690 CAGGAACCTGAGTGAGGCCCTCTGCTGCTCAGGATGGTCTCGGGGCGCTCGTGT 2749  
Qy 1790 CATCAAGACACCCCTCTGCCAGCTCACACATCTTTCATCACGACAAACGCCAGGACTT 1849  
Db 2750 CATCAAGACACCCCTCTGCCAGCTCACACATCTTTCATCACGACAAACGCCAGGACTT 2809  
Qy 1850 GGCTCCCCATCTCAGAACTCACAAGCATTTGCTCCCCAGCTGGGAGAACCTCAACCTCC 1909  
Db 2810 GGCTCCCCATCTCAGAACTCACAAGCATTTGCTCCCCAGCTGGGAGAACCTCAACCTCC 2869  
Qy 1910 CCCCTGCTCGTGGTGACAGAGGGGTGGACAGGGGGTGGGGGCTTCCCTCTGTACAT 1969  
Db 2870 CCCCTGCTCGTGGTGACAGAGGGGTGGACAGGGGGTGGGGGCTTCCCTCTGTACAT 2929  
Qy 1970 ACCCTGCCATACCAACCCAGGTATTAAATCTCGCTGGTGTCTTTTATTTTATTTT 2029  
Db 2930 ACCCTGCCATACCAACCCAGGTATTAAATCTCGCTGGTGTCTTTTATTTTATTTT 2989  
Qy 2030 TTGTTTGGATTTTAAATAAGAAATTTTCATTTTAAGCAAAAAA 2076  
Db 2990 TTGTTTGGATTTTAAATAAGAAATTTTCATTTTAAGCAAAAAA 3036  
  
RESULT 6  
AX333042 2907 bp DNA linear PAT 09-JAN-2002  
LOCUS AX333042  
DEFINITION Sequence 3551 from Patent WO0194629.  
ACCESSION AX333042  
VERSION AX333042.1 GI:18123676  
KEYWORDS human.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1  
REFERENCE  
AUTHORS Young, P. E., Augustus M., Carter, K. C., Ebner, R., Endress, G.,  
Horrigan, S., Soppet, D. R. and Weaver, Z.  
TITLE Cancer gene determination and therapeutic screening using signature  
gene sets  
JOURNAL Patent: WO 0194629-A 3551 13-DEC-2001;  
Avalon Pharmaceuticals (US)  
FEATURES  
Location/Qualifiers  
1..2907  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
BASE COUNT 561 a 1011 c 789 g 546 t  
ORIGIN  
Query Match 85.3%; Score 1778.8; DB 6; Length 2907;  
Best Local Similarity 99.3%; Pred. No. 3e-312;  
Matches 1786; Conservative 0; Mismatches 12; Indels 0; Gaps 0;  
  
Qy 288 CTCATTGAGACCCAGCAGCAGCTTCTGAAGAGATAGTGCACCGCTCTCCCTGCCACC 347  
Db 279 CACCATTGAGACCCAGCAGCAGCTTCTGAAGAGATAGTGCACCGCTCTCCCTGCCACC 338  
Qy 348 CCCTCTACCCCGCATCTACAGGCTTGTCTGTCTGAGGAGTCTCTCAGGAGTCTCAGGCTACCA 407  
Db 339 CCCTCTACCCCGCATCTACAGGCTTGTCTGTCTGAGGAGTCTCTCAGGAGTCTCAGGCTACCA 398  
Qy 408 CTATGGGCTCAGCGCTGTGAGGGCTCAAGGGCTTCTTCGCGCGCAGCAGCTCCAGAGAA 467  
Db 399 CTATGGGCTCAGCGCTGTGAGGGCTCAAGGGCTTCTTCGCGCGCAGCAGCTCCAGAGAA 458  
Qy 468 CATGGGTATACAGTGTACCGGACAAAGAACTGCATCATCAACAGGTGACCCGGAACCC 527  
Db 459 CATGGGTATACAGTGTACCGGACAAAGAACTGCATCATCAACAGGTGACCCGGAACCC 518  
Qy 528 CTGCCAGTACTGCCGACTGCAGAAAGTCTTTGAAGTGGGATGTCCAAAGGAGTCTGTGAG 587

```
Db 519 CTGCCAGTACTGCCGACTGCAGAGTGTCTTTGAAGTGGGCAATGTCCACGAGTCTGTGAG 578
Qy 588 AAACCAACCAACACAGAGAAGAGAGGTGCCCCAGCCCGAGTGTCTGTGAGAGCTACAC 647
Db 579 AAACCAACCAACAAAGAGAAGAGAGGTGCCCCAGCCCGAGTGTCTGTGAGAGCTACAC 638
Qy 648 GCTGACGCCGGAGGTGGGGAGCTCATTCAGAGAGTGGCCAAAGCGCACCCAGGAACCTT 707
Db 639 GCTGACGCCGGAGGTGGGGAGCTCATTCAGAGAGTGGCCAAAGCGCACCCAGGAACCTT 698
Qy 708 CCCTGCCCTCTGCCAGCTGGGCAATACACTACGAACAAACAGCTCAGAACAAACGTGTCTC 767
Db 699 CCCTGCCCTCTGCCAGCTGGGCAATACACTACGAACAAACAGCTCAGAACAAACGTGTCTC 758
Qy 768 TCTGGACATTGACCTCTGGGACAAAGTTCAAGTAACTCTCCACCAAGTGCATCATTAAAGC 827
Db 759 TCTGGACATTGACCTCTGGGACAAAGTTCAAGTAACTCTCCACCAAGTGCATCATTAAAGC 818
Qy 828 TGTGGAGTTCCCAAGCAGCTGCCGGCTTCAACACCTCACCACCTCACCATCGCCGACGATCAC 887
Db 819 TGTGGAGTTCCCAAGCAGCTGCCGGCTTCAACACCTCACCACCTCACCATCGCCGACGATCAC 878
Qy 888 CCTCCTCAAGGCTGCCTGCTGAGATCCTGTATCTGCGGATCTGCACGCGGTACACGCC 947
Db 879 CCTCCTCAAGGCTGCCTGCTGAGATCCTGTATCTGCGGATCTGCACGCGGTACACGCC 938
Qy 948 CGAGCAGACACCAATGACCTTCTCGGAGGGCTGACCTGAAACCGGACCCAGATGCACAA 1007
Db 939 CGAGCAGACACCAATGACCTTCTCGGAGGGCTGACCTGAAACCGGACCCAGATGCACAA 998
Qy 1008 CGCTGGCTTCCGCCCTTCCAGCAGCTGTCTTTGGCTTGCCCAACCAAGCTGCTGCCCT 1067
Db 999 CGCTGGCTTCCGCCCTTCCAGCAGCTGTCTTTGGCTTGCCCAACCAAGCTGCTGCCCT 1058
Qy 1068 GGAGATGGATGATCGGAGAGGGGCTGCTCAGGCGCATCTGCGCTCATCTCGGAGACCG 1127
Db 1059 GGAGATGGATGATCGGAGAGGGGCTGCTCAGGCGCATCTGCGCTCATCTCGGAGACCG 1118
Qy 1128 CCAGGACCTGGAGACCGCGGACGGGTGGACATGCTCAGGAGCGCTGCTGGAGGGCT 1187
Db 1119 CCAGGACCTGGAGACCGCGGACGGGTGGACATGCTCAGGAGCGCTGCTGGAGGGCT 1178
Qy 1188 AAGGTCTAGTGGGAGCGGAGGCCAGCCGCCCCACATGTTCCCAAGATGCTAAT 1247
Db 1179 AAGGTCTAGTGGGAGCGGAGGCCAGCCGCCCCACATGTTCCCAAGATGCTAAT 1238
Qy 1248 GAAGATTACTGACCTGCGAAGCATCAGCCGCAAGGGGCTGAGCGGTGATCAGCTGAA 1307
Db 1239 GAAGATTACTGACCTGCGAAGCATCAGCCGCAAGGGGCTGAGCGGTGATCAGCTGAA 1298
Qy 1308 GATGGAGATCCCGGCTCCATGCGGCTCTCATCCAGGAATGTTGGAGAACTCAGAGG 1367
Db 1299 GATGGAGATCCCGGCTCCATGCGGCTCTCATCCAGGAATGTTGGAGAACTCAGAGG 1358
Qy 1368 CTTGGACATCTGAGCGGACAGCCGGGGGTGGGGGCGGAGCGGGGTGGCTGCCCTCC 1427
Db 1359 CTTGGACATCTGAGCGGACAGCCGGGGGTGGGGGCGGAGCGGGGTGGCTGCCCTCC 1418
Qy 1428 CCGCCAGGACGCTGAGCCCGGAGCTCAGCCCGGAGCTCCACAGAGAGAGCCCGGAC 1487
Db 1419 CCGCCAGGACGCTGAGCCCGGAGCTCAGCCCGGAGCTCCACAGAGAGAGCCCGGAC 1478
Qy 1488 CCACCTCCCGTACCGCCCGGAGCTGACAGACGCTCGCCCTCGCCCTCGCCCGGCTTTT 1547
Db 1479 CCACCTCCCGTACCGCCCGGAGCTGACAGACGCTCGCCCTCGCCCTCGCCCGGCTTTT 1538
Qy 1548 CTCTGCTTTTACCGACCATGTACCCCGGACCGACCGCTCGCCCTCGCCCTCGCCCG 1607
Db 1539 CTCTGCTTTTACCGACCATGTACCCCGGACCGACCGCTCGCCCTCGCCCTCGCCCG 1598
Qy 1608 CGAGTACTGGGACCTTCCCTGGGAGCGGGGAGGAGGAGGAGGAGCTCCTTTGGACAG 1667
Db 1599 CGAGTACTGGGACCTTCCCTGGGAGCGGGGAGGAGGAGGAGGAGGAGGAGCTCCTTTGGACAG 1658
```

```
Qy 1668 AGGCCTGGGCCCTCACTGGACTGCCTGCCACAGCCTGGGCTGAGCTCAGAGGCCGAG 1727
Db 1659 AGGCCTGGGCCCTCACTGGACTGCCTGCCACAGCCTGGGCTGAGCTCAGAGGCCGAG 1718
Qy 1728 GCCAGAACTGAGTGAAGGCCCTGGTCCCTGAGGATGGTCTCCTGGGGCTCGTG 1787
Db 1719 GCCAGAACTGAGTGAAGGCCCTGGTCCCTGAGGATGGTCTCCTGGGGCTCGTG 1778
Qy 1788 TTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAACGCCAGGAC 1847
Db 1779 TTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTCATCACCAGCAACGCCAGGAC 1838
Qy 1848 TTGGCTCCGCCATCTCAGAACTCACAAGCCATTTGCTCCAGCTGGGAACTCAACCT 1907
Db 1839 TTGGCTCCGCCATCTCAGAACTCACAAGCCATTTGCTCCAGCTGGGAACTCAACCT 1898
Qy 1908 CCCCCCTGCCTGGTGTGACAGAGGGGTGGGACAGGGGGGGTTCGCCCTGTAC 1967
Db 1899 CCCCCCTGCCTGGTGTGACAGAGGGGTGGGACAGGGGGGGTTCGCCCTGTAC 1958
Qy 1968 ATACCTGCCATACCAACCCAGGTATTAAATTCCTGCTGGTGTGTTTAAATTT 2027
Db 1959 ATACCTGCCATACCAACCCAGGTATTAAATTCCTGCTGGTGTGTTTAAATTT 2018
Qy 2028 TTTTCTTTTGAATTTTAAATTAAGAAATTTTCAATTTAAGCACAAAAA 2085
Db 2019 TTTTCTTTTGAATTTTAAATTAAGAAATTTTCAATTTAAGCACATTATATCTGAAGGAA 2076
```

RESULT 7  
HSRRA  
LOCUS

DEFINITION Human mRNA for receptor of retinoic acid.  
ACCESSION X06614  
VERSION X06614.1 GI:36156  
KEYWORDS hormone receptor; retinoic acid receptor.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 2907)  
AUTHORS Giguere,V., Ong,E.S., Segui,P. and Evans,R.M.  
TITLE Identification of a receptor for the morphogen retinoic acid  
JOURNAL Nature 330 (6149), 624-629 (1987)  
MEDLINE 88065922  
PUBMED 2825036  
REFERENCE 2 (bases 1 to 2907)  
AUTHORS Banihmad,A., Kohne,A.C. and Renkawitz,R.  
TITLE A transferable silencing domain is present in the thyroid hormone receptor, in the v-erbA oncogene product and in the retinoic acid receptor  
JOURNAL EMBO J. 11 (3), 1015-1023 (1992)  
MEDLINE 92191988  
PUBMED 1347744  
FEATURES  
Location/Qualifiers  
1..2907  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="lambda hK1R"  
/clone\_lib="lambda gt10"  
103..1491  
/note="put. retinoic acid receptor (AA 1-462)"  
/codon\_start=1  
/protein\_id="CAA29829.1"  
/db\_xref="GI:36157"  
/db\_xref="SWISS-PROT:P10276"

CDS

1119 CCAGGACCTGGAGCAGCCGGACCGGGTGGACATGCTGCAGGAGCCGCTGCTGGAGGCGCT 1178

UNCLASSIFIED:  
1 (bases 1 to 2928)  
REFERENCE

Unclassified.





Thompson, C.C.  
TITLE Retinoic acid receptor composition and method for identifying  
ligands  
JOURNAL Patent: EP 0325849-A2 1 02-AUG-1989;  
FEATURES Location/Qualifiers  
source 1. 2908  
BASE COUNT 562 a 1012 c 788 g 546 t  
ORIGIN

Query Match 85.2%; Score 1777.2; DB 6; Length 2908;  
Best Local Similarity 99.3%; Pred. No. 6e-312;  
Matches 1785; Conservative 0; Mismatches 13; Indels 0; Gaps 0;  
QY 288 CTCATTGAGACCCAGACGACAGTTCGAAAGAGATAGTGCCAGCCCTCCCTGCCACC 347  
DB 279 CACCATTTAGAGACCCAGACGACAGTTCGAAAGAGATAGTGCCAGCCCTCCCTGCCACC 338  
QY 348 CCCTCTACCCCGCATCTACAAGCCTTGTCTGTCTGAGGACAAAGTCTCAGGCTACCA 407  
DB 339 CCCTCTACCCCGCATCTACAAGCCTTGTCTGTCTGAGGACAAAGTCTCAGGCTACCA 398  
QY 408 CTATGGGTGACGCGCTGTGAGGGTGCAGAGGCTTCTTCGCGCCGACGATCCAGAGAA 467  
DB 399 CTATGGGTGACGCGCTGTGAGGGTGCAGAGGCTTCTTCGCGCCGACGATCCAGAGAA 458  
QY 468 CATGGTGTACAGCTGTACCCGGGACAAAGTGTATCATCAACAAAGTGTACCCGGAACCC 527  
DB 459 CATGGTGTACAGCTGTACCCGGGACAAAGTGTATCATCAACAAAGTGTACCCGGAACCC 518  
QY 528 CTGCCAGTACTGCCAGTGCAGAGTGTCTTGAAGTGGCATGTCCAAGGAGTCTGTGAG 587  
DB 519 CTGCCAGTACTGCCAGTGCAGAGTGTCTTGAAGTGGCATGTCCAAGGAGTCTGTGAG 578  
QY 588 AAACGACCGAAACAAAGAGAGAGAGTGCACCAAGCCGAGTCTCTGAGAGCTACAC 647  
DB 579 AAACGACCGAAACAAAGAGAGAGAGTGCACCAAGCCGAGTCTCTGAGAGCTACAC 638  
QY 648 GCTGACGCGGAGGTGGGGAGCTCATTTAGAAGTGGCGAAAGCCACGAGAAACCTT 707  
DB 639 GCTGACGCGGAGGTGGGGAGCTCATTTAGAAGTGGCGAAAGCCACGAGAAACCTT 698  
QY 708 CCCTGCCCTCTGCCAGTGGGCAATATACACTAGCAACACAGCTCAGAACACGTGTCTC 767  
DB 699 CCCTGCCCTCTGCCAGTGGGCAATATACACTAGCAACACAGCTCAGAACACGTGTCTC 758  
QY 768 TCTGACATTCACCTCTCGGACAAAGTTCAGTGAACCTTCCACCAAGTGCATCATTAAGAC 827  
DB 759 TCTGACATTCACCTCTCGGACAAAGTTCAGTGAACCTTCCACCAAGTGCATCATTAAGAC 818  
QY 828 TGTGAGTTCGCCAAGCAGCTGCCGGCTTCACCAACCTCACCATCGCCGACGAGATCAC 887  
DB 819 TGTGAGTTCGCCAAGCAGCTGCCGGCTTCACCAACCTCACCATCGCCGACGAGATCAC 878  
QY 888 CCTCTCAAGGCTGCTGCCCTGGACATCCTGATCTGGGATCTGACGCGGTACACGCC 947  
DB 879 CCTCTCAAGGCTGCTGCCCTGGACATCCTGATCTGGGATCTGACGCGGTACACGCC 938  
QY 948 CGAGCAGACACCATGACCTTCTCGGAGGGCTGACCTGACCTGACCGGACCCAGATGCACAA 1007  
DB 939 CGAGCAGACACCATGACCTTCTCGGAGGGCTGACCTGACCTGACCGGACCCAGATGCACAA 998  
QY 1008 CGCTGGCTTCGGCCCTCCACGACCTGTCTTTCCTTTCGCAACACGAGCTGCTGCCCT 1067  
DB 999 CGCTGGCTTCGGCCCTCCACGACCTGTCTTTCCTTTCGCAACACGAGCTGCTGCCCT 1058  
QY 1068 GGAGATGGATGATCGGAGACGGGCTGCTCAGCGCCATCTGCCCTATCTGCGGAGACCG 1127  
DB 1059 GGAGATGGATGATCGGAGACGGGCTGCTCAGCGCCATCTGCCCTATCTGCGGAGACCG 1118  
QY 1128 CCAGACCTGAGACGCGGACCGGCTGGACATGCTGAGGAGCGGCTGCTGGAGCGCT 1187  
DB 1119 CCAGACCTGAGACGCGGACCGGCTGGACATGCTGAGGAGCGGCTGCTGGAGCGCT 1178

QY 1188 AAAGCTCTACGTGCGGAAGCGGAGCGCCAGCGCCACATGTTCCCAAGATGCTAAT 1247  
DB 1179 AAAGGTCTACGTGCGGAAGCGGAGCGCCAGCGCCACATGTTCCCAAGATGCTAAT 1238  
QY 1248 GAAGATTACTGACCTGCGGAAGCATCAGCGCCAAAGGGGGTGTAGCGGGTGTATCAGCTGAA 1307  
DB 1239 GAAGATTACTGACCTGCGGAAGCATCAGCGCCAAAGGGGGTGTAGCGGGTGTATCAGCTGAA 1298  
QY 1308 GATGAGATCCCGGGTTCATGCCGCTCTCATCAGGAAATTTTGGAGAACTCAGAGG 1367  
DB 1299 GATGAGATCCCGGGTTCATGCCGCTCTCATCAGGAAATTTTGGAGAACTCAGAGG 1358  
QY 1368 CCTGACACTCTGAGCGGACACCGGGGGTGTGGGGGGGAGCGGGGTGTGCGCTGCCCC 1427  
DB 1359 CCTGACACTCTGAGCGGACACCGGGGGTGTGGGGGGGAGCGGGGTGTGCGCTGCCCC 1418  
QY 1428 CCGCGCAGCAGCTGTAGCCCGCAGCTCAGCCCCAGCTCACAAGAGCAGCCCGCCAC 1487  
DB 1419 CCGCGCAGCAGCTGTAGCCCGCAGCTCAGCCCCAGCTCACAAGAGCAGCCCGCCAC 1478  
QY 1488 CCACCTCCCGGTGACCGCCCGCAGCCCATGGACACAGCCCTCGCCCTCGCCCGCGCTTTT 1547  
DB 1479 CCACCTCCCGGTGACCGCCCGCAGCCCATGGACACAGCCCTCGCCCTCGCCCGCGCTTTT 1538  
QY 1548 CTCTGCTTTTACCGACCATGTGACCCCGCAGCCAGCCCTCGCCCGCGCTTTT 1607  
DB 1539 CTCTGCTTTTACCGACCATGTGACCCCGCAGCCAGCCCTCGCCCGCGCTTTT 1598  
QY 1608 GCAGTACTGGGACCTTCCCTGGGGACGGGAGGAGGAGGAGGAGCTCTTGGACAG 1667  
DB 1599 GCAGTACTGGGACCTTCCCTGGGGACGGGAGGAGGAGGAGGAGGAGCTCTTGGACAG 1658  
QY 1668 AGGCTTGGGCGCTCAGTGGACTGCTTCCACAGCCTGGGTGACGTGAGGCGCGAG 1727  
DB 1659 AGGCTTGGGCGCTCAGTGGACTGCTTCCACAGCCTGGGTGACGTGAGGCGCGAG 1718  
QY 1728 GCCAGAACTGAGTGGGCGCTGCTGGGTCTCAGGATGGGTCTGGGGCGCTCGTG 1787  
DB 1719 GCCAGAACTGAGTGGGCGCTGCTGGGTCTCAGGATGGGTCTGGGGCGCTCGTG 1778  
QY 1788 TTTCATCAAGACACCTCTGCCAGCTCAGCAGCTTTCATCACCAGCAACGCCAGGAC 1847  
DB 1779 TTTCATCAAGACACCTCTGCCAGCTCAGCAGCTTTCATCACCAGCAACGCCAGGAC 1838  
QY 1848 TTGGTTCGCCCATCTCAGAACTCACAAGCCTTGTCTCCACAGCTGGGGAACCTCAACCT 1907  
DB 1839 TTGGTTCGCCCATCTCAGAACTCACAAGCCTTGTCTCCACAGCTGGGGAACCTCAACCT 1898  
QY 1908 CCCCCCTGCTCGGTGGTGTGACAGAGGGGTGGGACAGGGGGGGGGTTCGCCCTGTAC 1967  
DB 1899 CCCCCCTGCTCGGTGGTGTGACAGAGGGGTGGGACAGGGGGGGGGTTCGCCCTGTAC 1958  
QY 1968 ATACCTCGCATACCAACCCAGGTATTATCTCGCTGGTTTGTGTTTATTTAATTT 2027  
DB 1959 ATACCTCGCATACCAACCCAGGTATTATCTCGCTGGTTTGTGTTTATTTAATTT 2018  
QY 2028 TTTTGTGTTGATTTTTTAAAGAAATTTTCATTTTAAAGCAAAAAA 2085  
DB 2019 TTTTGTGTTGATTTTTTAAAGAAATTTTCATTTTAAAGCAAAAAA 2076

RESULT 10  
LOCUS AR052136 2940 bp DNA linear PAT 29-SEP-1999  
DEFINITION Sequence 1 from patent US 5830760.  
ACCESSION AR052136  
VERSION AR052136.1 GI:5975500  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 2940)



AUTHORS Tsai, S. and Collins, S.J.  
TITLE Creating novel hematopoietic cell lines by expressing altered retinoic acid receptors  
JOURNAL Patent: US 5830760-A 1 03-NOV-1998;  
FEATURES Location/Qualifiers  
source 1. 2940  
BASE COUNT 594 a 1012 c 788 g 546 t  
ORIGIN

Query Match 85.2%; Score 1777.2; DB 6; Length 2940;  
Best Local Similarity 99.3%; Pred. No. 6e-312;  
Matches 1785; Conservative 0; Mismatches 13; Indels 0; Gaps 0;

QY 288 CTCATTGAGACCCAGACAGCAGTCTGAGAGAGATAGTGGCCAGCCCTCCCTCCGCCACC 347  
DB 279 CACCATTTGAGACCCAGACAGCAGTCTGAGAGAGATAGTGGCCAGCCCTCCCTCCGCCACC 338  
QY 348 CCCTCTACCCCGCATCTACAAAGCCTTGCTTTGCTGTCTGAGGACAAAGTCTCTCAGGCTACCA 407  
DB 339 CCCTCTACCCCGCATCTACAAAGCCTTGCTTTGCTGTCTGAGGACAAAGTCTCTCAGGCTACCA 398  
QY 408 CTATGGGGTCAGCCGCTGTGAGGGCTGCAAGGGCTTTCTCCGCCAGCATCCAGGAAGAA 467  
DB 399 CTATGGGGTCAGCCGCTGTGAGGGCTGCAAGGGCTTTCTCCGCCAGCATCCAGGAAGAA 458  
QY 468 CATGGTCTACAGCTGTACCCGGGACAAAGTGTGATCATCAACAGGTGACCCGGAACCC 527  
DB 459 CATGGTCTACAGCTGTACCCGGGACAAAGTGTGATCATCAACAGGTGACCCGGAACCC 518  
QY 528 CTGCCAGTACTCGCGACTGCAAGAGTGTCTTGAAGTGGGCGATGTCCAAAGGAGTCTGTGAG 587  
DB 519 CTGCCAGTACTCGCGACTGCAAGAGTGTCTTGAAGTGGGCGATGTCCAAAGGAGTCTGTGAG 578  
QY 588 AAACGACCGAAACAAAGAAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTGTGAGAGTACAC 647  
DB 579 AAACGACCGAAACAAAGAAAGAAAGAGAGGTGCCCAAGCCCGAGTGTCTGTGAGAGTACAC 638  
QY 648 GCTCACCCGAGGTGGGGAGTCAATTGAGAAGGTGGCAAGCGCACCGACCGAAGACCTT 707  
DB 639 GCTCACCCGAGGTGGGGAGTCAATTGAGAAGGTGGCAAGCGCACCGACCGAAGACCTT 698  
QY 708 CCCTGSCCCTCTGCCAGCTGGGCAATACACTACGAACAAACAGCTCAGAACACAGTGTCTC 767  
DB 699 CCCTGSCCCTCTGCCAGCTGGGCAATACACTACGAACAAACAGCTCAGAACACAGTGTCTC 758  
QY 768 TCTGGACATTTGACCTCTGGGCAAGTTCAGTGAATCTTCCACCAAGTGCATTAAGAC 827  
DB 759 TCTGGACATTTGACCTCTGGGCAAGTTCAGTGAATCTTCCACCAAGTGCATTAAGAC 818  
QY 828 TGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAACACCCCTACCATCGCCGACCATCAC 887  
DB 819 TGTGGAGTTCGCCAAGCAGCTGCCGGCTTCAACACCCCTACCATCGCCGACCATCAC 878  
QY 888 CCTCCTCAAGGCTGCCCTGGACATCTCTGATCTCGGGATCTGCACGCGGTACACGCC 947  
DB 879 CCTCCTCAAGGCTGCCCTGGACATCTCTGATCTCGGGATCTGCACGCGGTACACGCC 938  
QY 948 CGAGCAGACACCATGACCTTCTCGGAGGGCTGACCCCTGAACCGGACCCAGATGACAA 1007  
DB 939 CGAGCAGACACCATGACCTTCTCGGAGGGCTGACCCCTGAACCGGACCCAGATGACAA 998  
QY 1008 CGCTGGCTTCGGCCCTTCACCGACCTGGTCTTTGGCTTCGCCCAACAGCTGTGCCCCCT 1067  
DB 999 CGCTGGCTTCGGCCCTTCACCGACCTGGTCTTTGGCTTCGCCCAACAGCTGTGCCCCCT 1058  
QY 1068 GGAGATGGATGATCGGAGAGCGGCTGCTCAGCGCCATCTGCTCATCTCGGAGACCG 1127  
DB 1059 GGAGATGGATGATCGGAGAGCGGCTGCTCAGCGCCATCTGCTCATCTCGGAGACCG 1118  
QY 1128 CCAGGACCTGGAGCAGCGGACCGGCTGGACATGCTGCAGGAGCGGCTGTCTGGAGGCGCT 1187  
DB 1119 CCAGGACCTGGAGCAGCGGACCGGCTGGACATGCTGCAGGAGCGGCTGTCTGGAGGCGCT 1178

QY 1188 AAAGTCTACGTGCGGAACCGGAGGCCAGCCGCCACATGTTCCCAAGATGCTAAT 1247  
DB 1179 AAAGTCTACGTGCGGAACCGGAGGCCAGCCGCCACATGTTCCCAAGATGCTAAT 1238  
QY 1248 GAAGATTACTGACCTGCGAAGCATCAGCCAGGGGGTGTAGCGGTGTATCAGCTGAA 1307  
DB 1239 GAAGATTACTGACCTGCGAAGCATCAGCCAGGGGGTGTAGCGGTGTATCAGCTGAA 1298  
QY 1308 GATGAGATTCGCGGGCTCCATGCCGCTCTCATPCAGGAAATTTGGAGAACTCAGAGGG 1367  
DB 1299 GATGAGATTCGCGGGCTCCATGCCGCTCTCATPCAGGAAATTTGGAGAACTCAGAGGG 1358  
QY 1368 CTTGGACACTCTGAGGGACACCGGGGGTGGGGGGCGGGGGTGGCGCTGCCCC 1427  
DB 1359 CTTGGACACTCTGAGCGGACACCGGGGGTGGGGGGCGGGGGTGGCGCTGCCCC 1418  
QY 1428 CCGCGCAGCAGCTGTAGCCCGCAGCTCAGCCCGCAGCTCAACAGAGCAGCCCGGCCAC 1487  
DB 1419 CCGCGCAGCAGCTGTAGCCCGCAGCTCAGCCCGCAGCTCAACAGAGCAGCCCGGCCAC 1478  
QY 1488 CCACCTCCCGGTGACCGCCACCGCCACATGGACACACGCTCGCCCTCCCGCCCGGCTTTT 1547  
DB 1479 CCACCTCCCGGTGACCGCCCGCCACATGGACACACGCTCGCCCTCCCGCCCGGCTTTT 1538  
QY 1548 CTCCTGCTTTTACCGACCATGTGACCCCGCAGCCCTGCCCCACCTGCCCTCCCGG 1607  
DB 1539 CTCCTGCTTTTACCGACCATGTGACCCCGCAGCCCTGCCCCACCTGCCCTCCCGG 1598  
QY 1608 GCAGTACTGGGACCTTTCCTGGGGACCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1667  
DB 1599 GCAGTACTGGGACCTTTCCTGGGGACCGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1658  
QY 1668 AGGCTGGGCGCTCAGTGGACTGCTGCTCCACAGCTGGGTGAGTGTGAGGAGGAGGAG 1727  
DB 1659 AGGCTGGGCGCTCAGTGGACTGCTGCTCCACAGCTGGGTGAGTGTGAGGAGGAGGAG 1718  
QY 1728 GCAGGAACTGAGTGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1787  
DB 1719 GCAGGAACTGAGTGGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1778  
QY 1788 TTTCATCAAGACACCCCTCTGCCCCAGCTCACCACATCTTCATCACCAGCAACCCAGGAC 1847  
DB 1779 TTTCATCAAGACACCCCTCTGCCCCAGCTCACCACATCTTCATCACCAGCAACCCAGGAC 1838  
QY 1848 TTGGCTCCCGCATCTCAGAACTCAGACCCATGTCTCCCGAGCTGGGAGACCTCAACCT 1907  
DB 1839 TTGGCTCCCGCATCTCAGAACTCAGACCCATGTCTCCCGAGCTGGGAGACCTCAACCT 1898  
QY 1908 CCCCCCTGCTCGGTGGTGACAGAGGGGTGGGACAGGGGGGGGGTTCGCCCTGTAC 1967  
DB 1899 CCCCCCTGCTCGGTGGTGACAGAGGGGTGGGACAGGGGGGGGGTTCGCCCTGTAC 1958  
QY 1968 ATACCTGCTCATACCAACCCAGGATTAATTTCTCGCTGGTTTGTGTTTATTTAATTT 2027  
DB 1959 ATACCTGCTCATACCAACCCAGGATTAATTTCTCGCTGGTTTGTGTTTATTTAATTT 2018  
QY 2028 TTTTCTTTTGTATTTTAAATAAGAAATTTTCAATTTTAAAGCAAAAAAATAAAAAA 2085  
DB 2019 TTTTCTTTTGTATTTTAAATAAGAAATTTTCAATTTTAAAGCAACATTTTAACTGAAGAA 2076

RESULT 11  
109348  
LOCUS I09348 Sequence 1 from Patent WO 8905355.  
ACCESSION I09348  
VERSION I09348.1 GI:587943  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 2940)  
DNA linear PAT 02-DEC-1994

AUTHORS		Evans, R.M., Giguere, V., Ong, E.S., Segui, P.S., Umesono, K. and Thompson, C.C.									
JOURNAL		Patent: WO 8905355-A 1 15-JUN-1989;									
FEATURES		Location/Qualifiers									
source		1..2940									
BASE COUNT		/organism="unknown"									
ORIGIN		594 a 1013 c 787 g 546 t									
Query Match		85.2%; Score 1777.2; DB 6; Length 2940;									
Best Local Similarity		99.3%; Pred. No. 66-312;									
Matches 1785; Conservative		0; Mismatches 13; Indels 0; Gaps 0;									
QY	288	CTCCATTGAGACCCAGAGCAGTTCGAGAGATAGTGGCCAGCCCTCCCTCGCCACC	347								
Db	279	CACCATTTAGACCCAGACGACGAGTCTTGAGAGATAGTGGCCAGCCCTCCCTCGCCACC	338								
QY	348	CCCTCTACCCCGCATCTACAAGCCCTTGCTTGTCTGTCTCAGGACAAAGTCTCAGGGTACCA	407								
Db	339	CCCTCTACCCCGCATCTACAAGCCCTTGCTTGTCTGTCTCAGGACAAAGTCTCAGGGTACCA	398								
QY	408	CTATGGGGTTCAGCGCTGTGAGGGTTCGAAGGGCTTCTTCCGCCGAGCATCCAGAAGAA	467								
Db	399	CTATGGGGTTCAGCGCTGTGAGGGTTCGAAGGGCTTCTTCCGCCGAGCATCCAGAAGAA	458								
QY	468	CATGGTGTACAGTGTCTACCCGGGACAAAGTGCATCATCAACAAGGTGACCCGGAAACCC	527								
Db	459	CATGGTGTACAGTGTCTACCCGGGACAAAGTGCATCATCAACAAGGTGACCCGGAAACCG	518								
QY	528	CTGCCAGTACTGCCGACTGCAGAACTGCTTGAAGTGGGATGTCCAAAGAGTCTGTGAG	587								
Db	519	CTGCCAGTACTGCCGACTGCAGAACTGCTTGAAGTGGGATGTCCAAAGAGTCTGTGAG	578								
QY	588	AAACGACCGAAACAAGAAGAAGAGGTGCCCAAGCCGAGTCTCTGAGAGCTACAC	647								
Db	579	AAACGACCGAAACAAGAAGAAGAGGTGCCCAAGCCGAGTCTCTGAGAGCTACAC	638								
QY	648	GCTGACCGGAGGTGGGGAGCTCATTTGAGNAGTGGCGCAAGCGCACCAGGAACCTT	707								
Db	639	GCTGACCGGAGGTGGGGAGCTCATTTGAGNAGTGGCGCAAGCGCACCAGGAACCTT	698								
QY	708	CCCTGCCCTCTGCCAGCTGGGCAATACACTACGAAACAACAGCTCAGAAACAAGTGTCTC	767								
Db	699	CCCTGCCCTCTGCCAGCTGGGCAATACACTACGAAACAACAGCTCAGAAACAAGTGTCTC	758								
QY	768	TCTGACATTTGACCTCTGGGCAAGTTCAGTGAAGTCTTCCACCAAGTGCATCATTAAGAC	827								
Db	759	TCTGACATTTGACCTCTGGGCAAGTTCAGTGAAGTCTTCCACCAAGTGCATCATTAAGAC	818								
QY	828	TGTGAGTTTCCCAAGCAGCTGCCGGCTTACACACCTCACCATCGCCGACCATCAC	887								
Db	819	TGTGAGTTTCCCAAGCAGCTGCCGGCTTACACACCTCACCATCGCCGACCATCAC	878								
QY	888	CCTCCTCAAGGTTGCTGCTGGACATCTGTATCTCTGGGATCTGCACGGGTACAGCC	947								
Db	879	CCTCCTCAAGGTTGCTGCTGGACATCTGTATCTCTGGGATCTGCACGGGTACAGCC	938								
QY	948	CGAGCAGGACACCATGACCTTCTCGGACGGGTGACCTGAAACCGGACCCAGATGCACAA	1007								
Db	939	CGAGCAGGACACCATGACCTTCTCGGACGGGTGACCTGAAACCGGACCCAGATGCACAA	998								
QY	1008	CGCTGGCTTCGGCCCTTCACGACCTGGTCTTTTGCCTTCGCCCAACCAAGCTGTGCCCT	1067								
Db	999	CGCTGGCTTCGGCCCTTCACGACCTGGTCTTTTGCCTTCGCCCAACCAAGCTGTGCCCT	1058								
QY	1068	GGAGATGGATGATGCGGAGACGGGGTGTCTCAGCGCCATCTGCTCATCTCGGAGACCG	1127								
Db	1059	GGAGATGGATGATGCGGAGACGGGGTGTCTCAGCGCCATCTGCTCATCTCGGAGACCG	1118								
QY	1128	CAAGGACCTGGAGCAGCCGGAGCGGGTGGACATGCTGAGAGCGCGTGTGTGAGGCGCT	1187								
Db	1119	CAAGGACCTGGAGCAGCCGGAGCGGGTGGACATGCTGAGAGCGCGTGTGTGAGGCGCT	1178								

QY	1188	AAAGTCTACTGTCGCGAAGCGAGGCCGCCAGCCGCCACATGTTCCCAAGATGCTAAT	1247								
DB	1179	AAAGTCTACTGTCGCGAAGCGAGGCCGCCAGCCGCCACATGTTCCCAAGATGCTAAT	1238								
QY	1248	GAAGATTACTGACCTGCGAAGCATATCAGCCAAAGGGGCTGAGGGGTGATCAGCTGAA	1307								
DB	1239	GAAGATTACTGACCTGCGAAGCATATCAGCCAAAGGGGCTGAGGGGTGATCAGCTGAA	1298								
QY	1308	GATGGAGATCCCGGCTCCATGCCCTCTCATCCAGGAATGTTGAGAACTCAGAGGG	1367								
DB	1299	GATGGAGATCCCGGCTCCATGCCCTCTCATCCAGGAATGTTGAGAACTCAGAGGG	1358								
QY	1368	CTGGACACTCTCAGCGGACAGCGGGGGTGGGGGCGGAGCGGGGTGGCTGCCCCC	1427								
DB	1359	CTGGACACTCTGAGCGGACAGCGGGGGTGGGGGCGGAGCGGGGTGGCTGCCCCC	1418								
QY	1428	CCCGCCAGCAGCTGTAGCCGCCAGCTCAGCCAGTCCCAAGAGAGCGCGGCGAC	1487								
DB	1419	CCCGCCAGCAGCTGTAGCCGCCAGCTCAGCCAGTCCCAAGAGAGCGCGGCGAC	1478								
QY	1488	CCACTCCCGGTGACCGCCGACGACATGACAGCCCTCGCCCTCGCCCGGCTTTT	1547								
DB	1479	CCACTCCCGGTGACCGCCGACGACATGACAGCCCTCGCCCTCGCCCGGCTTTT	1538								
QY	1548	CTCTGCTTTTCTACCGACCATGTGACCCGACAGCCCTGCCCCACCTGCCCTCCCGG	1607								
DB	1539	CTCTGCTTTTCTACCGACCATGTGACCCGACAGCCCTGCCCCACCTGCCCTCCCGG	1598								
QY	1608	GCAGTACTGGGACCTTCCCTGGGGACGCGGGAGGAGGAGGAGGAGGAGGAGGAG	1667								
DB	1599	GCAGTACTGGGACCTTCCCTGGGGACGCGGGAGGAGGAGGAGGAGGAGGAGGAG	1658								
QY	1668	AGGCTGGGCGCTCAGTGGACCTGCTGCCACAGCTGGGCTGGGCTGAGGAGGAGG	1727								
DB	1659	AGGCTGGGCGCTCAGTGGACCTGCTGCCACAGCTGGGCTGGGCTGAGGAGGAGG	1718								
QY	1728	GCAGGAATGAGTGGGCGCTGCTGGGTCTCAGGATGGTCTCTGGGGCTCGTG	1787								
DB	1719	GCAGGAATGAGTGGGCGCTGCTGGGTCTCAGGATGGTCTCTGGGGCTCGTG	1778								
QY	1788	TTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTTCATCACCACCAAGCCAGGAC	1847								
DB	1779	TTTCATCAAGACACCCCTCTGCCAGCTCACCACATCTTTCATCACCACCAAGCCAGGAC	1838								
QY	1848	TTGGCTCCCGCATCTCAGAACTCAGAACTGCTCCCGAGCTGGGGGAGGAGGAGGAG	1907								
DB	1839	TTGGCTCCCGCATCTCAGAACTCAGAACTGCTCCCGAGCTGGGGGAGGAGGAGGAG	1898								
QY	1908	CCCGCTGCTCGGTTGGTGCAGAGGGGTGGGACAGGGGGGGGGGGTTCGCCCTGTAC	1967								
DB	1899	CCCGCTGCTCGGTTGGTGCAGAGGGGTGGGACAGGGGGGGGGGGTTCGCCCTGTAC	1958								
QY	1968	ATACCTGCTGCATACCAACCCAGGTATTAATCTCGCTGGTGGTGGTGGTGGTGGTGG	2027								
DB	1959	ATACCTGCTGCATACCAACCCAGGTATTAATCTCGCTGGTGGTGGTGGTGGTGGTGG	2018								
QY	2028	TTTTGTTTTGATTTTTTAAAGAAATTTTCATTTTAAGCACAATAAAAAA	2085								
DB	2019	TTTTGTTTTGATTTTTTAAAGAAATTTTCATTTTAAGCACAATAAAAAA	2076								

RESULT 12

AK098172

LOCUS

DEFINITION

AK098172

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

AK098172

2008 bp mRNA linear PRI 15-JUL-2002

Homo sapiens cDNA FLJ40853 fis, clone TRACH2015486, highly similar

to RETINOIC ACID RECEPTOR ALPHA.

GI:21758129

oligo capping; fis (full insert sequence).

Homo sapiens trachea cDNA to mRNA, clone\_lib:TRACH2

clone:TRACH2015486.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;



```
|||||
Db 1714 TGGGGGCTCGTGTTCATCAGACACCCCTCGCCAGCTCACACATCTTCATCACACAG 1773
Qy 1835 CAACGCCAGGACTTGGCTCCGCCATCTCAGAACTCACAAGCCATGCTCCCGAGCTGG 1894
Db 1774 CAACGCCAGGACTTGGCTCCGCCATCTCAGAACTCACAAGCCATGCTCCCGAGCTGG 1833
Qy 1895 GGAACCTCAACCTCCGCCCTCGCTGGTGTGTGACAGAGGGGTGGGACAGGGGGGGGG 1954
Db 1834 GGAACCTCAACCTCCGCCCTCGCTGGTGTGTGACAGAGGGGTGGGACAGGGGGGGGG 1893
Qy 1955 GTTCCCTCTGACATACCTCGCATACCAACCCAGGATTAATCTCGCTGGTGTGTGT 2014
Db 1894 GTTCCCTCTGACATACCTCGCATACCAACCCAGGATTAATCTCGCTGGTGTGTGT 1953
Qy 2015 TTTATTTAAATTTTGTGTTGATTTTAAATTAAGAAATTTTCAATTTAAGCAC 2069
Db 1954 TTTATTTAAATTTTGTGTTGATTTTAAATTAAGAAATTTTCAATTTAAGCAC 2008

RESULT 13
HSU41742
LOCUS
DEFINITION
NPW-RAR long form mRNA, complete cds.
ACCESSION
D41742
VERSION
D41742.1 GI:1314307
KEYWORDS
Homo sapiens.
SOURCE
Homo sapiens
ORGANISM
Homo sapiens
REFERENCE
1. HSD41742 2073 bp mRNA linear PRI 17-MAY-1996
2. Human nucleophosmin-retinoic acid receptor alpha fusion protein
NPW-RAR long form mRNA, complete cds.
AUTHORS
Redner, R.L., Rush, E.A., Faas, S., Rudert, W.A. and Corey, S.J.
TITLE
The t(5;17)(q32;q11) translocation
nucleophosmin-retinoic acid receptor fusion
JOURNAL
Blood 87 (3), 882-886 (1996)
MEDLINE
96151966
PUBMED
8562957
REFERENCE
2. (bases 1 to 2073)
Redner, R.L.
Direct Submission
TITLE
Submitted (30-NOV-1995) Robert L. Redner, Medicine, University of
Pittsburgh, E1058 BST, 211 Lothrop Street, Pittsburgh, PA 15213,
USA
FEATURES
Location/Qualifiers
1. 2073
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="acute promyelocytic leukemia cells with
t(5;17)(q32;q11) translocation"
/tissue_type="bone marrow"
84..1775
/codon_start=1
/product="nucleophosmin-retinoic acid receptor alpha
fusion protein NPW-RAR long form"
/protein_id="AAB00112.1"
/db_xref="GI:1314308"
/translation="MEDSMDMSPRLPQNYLFGCELKADKDYHPKVDNENEHOLS
RTVSLGAGAKDELHIVERAMNVEGSPILKVTATLKMVSQVPSLGGPEINPPVYLR
KCGSPVHSIQHVLVADEAEDEEEDENRVLFPKRGKLMHFGDLVCSKRLA
AIEFTSSSEIVSPSPPLPRYKPCFVCQDKSSGYHYGVSAECGCKGFRRSIQ
KNMYTCHRDKNKINRYNRQYCRLOKGFYVNGMSKESVRNDRKKKKEVPKPCS
ESYTLTPVGLIKRKAHQTFFALQKGFYTNNSSEORVSDIDLDWKFSELST
KCIKTVEFAKQLGFTTLTADQITLLKACLDLILIRICTRYTPEDDTWPFSLGT
LNRTOMHNAFGPLDVLVAFANQLLPLEMDAEGLISALICLGQRDLQEPQPRVD
MLQPLLEALKYVYVRKRPSPRPFPMKIDLRISAKGAERVITLKMIEGSP
PLIQEMLENSEGLDTLSGQPGGGGGGLAPPGPSLSLSPSSNRSSPATHSP"
BASE COUNT 465 a 612 c 605 g 391 t
ORIGIN
```

Query Match 71.9%; Score 1500.8; DB 9; Length 2073;  
Best Local Similarity 98.3%; Pred. No. 7.le-262;

```
Matches 1517; Conservative 0; Mismatches 27; Indels 0; Gaps 0;
Qy 255 GCTCCGGACTCCGCTTTGGAATGGCTCAACACCACTCCATTGAGACCCAGAGCAGCTTC 314
Db 530 GTTCGGGGATTGGTGTGAGTTGGAATTTGGCAGCCATTGAGACCCAGAGCAGCTTC 589
Qy 315 TGAAGAGATAGTGGCCAGCCCTCCCTCGCCACCCCTCTACCCCGCATCTACAAAGCCCTG 374
Db 590 TGAAGAGATAGTGGCCAGCCCTCCCTCGCCACCCCTCTACCCCGCATCTACAAAGCCCTG 649
Qy 375 CTTTGTCTCTCAGGACAACTCTCAGGCTACCACTATGCGGTGAGGCTCAGGCTGAGGCTG 434
Db 650 CTTTGTCTCTCAGGACAACTCTCAGGCTACCACTATGCGGTGAGGCTCAGGCTGAGGCTG 709
Qy 435 CAAGGCTCTCTCCGCCCGCAGCATCCAGAGAATGGTGTACACGTGTACCCGGGACAA 494
Db 710 CAAGGCTCTCTCCGCCCGCAGCATCCAGAGAATGGTGTACACGTGTACCCGGGACAA 769
Qy 495 GAACTGCATCATCAACAAGGTGACCCGGGACCCCTGCGAGTACTGCCAGCTGCAGAAAGTG 554
Db 770 GAACTGCATCATCAACAAGGTGACCCGGGACCCCTGCGAGTACTGCCAGCTGCAGAAAGTG 829
Qy 555 CTTTGAAGTGGGATCTCCAAGGAGTCTGTGAGAAACGACCCGAAACAAGAAAGAAAGGA 614
Db 830 CTTTGAAGTGGGATCTCCAAGGAGTCTGTGAGAAACGACCCGAAACAAGAAAGAAAGGA 889
Qy 615 GGTGCCAAAGCCGAGTGTCTGAGAGCTACACGCTGACCCGCGGAGGTGGGGGAGCTCAT 674
Db 890 GGTGCCAAAGCCGAGTGTCTGAGAGCTACACGCTGACCCGCGGAGGTGGGGGAGCTCAT 949
Qy 675 TGAGAGGTGGCCAAAGCCGACCCAGGAACCTTCCCTGCCCTCTGCCAGCTGGGCCAAATA 734
Db 950 TGAGAGGTGGCCAAAGCCGACCCAGGAACCTTCCCTGCCCTCTGCCAGCTGGGCCAAATA 1009
Qy 735 CACTACGAAACAACAGCTCAGAACAAAGTGTCTCTCTGGACATTTGACCTCTGGGACAAAGTT 794
Db 1010 CACTACGAAACAACAGCTCAGAACAAAGTGTCTCTCTGGACATTTGACCTCTGGGACAAAGTT 1069
Qy 795 CAGTGAACCTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 854
Db 1070 CAGTGAACCTCTCCACCAAGTGCATTAAGACTGTGGAGTTCGCCAAGCAGCTGCCCGG 1129
Qy 855 CTTACACCCCTCACCATCGCCAGCAGATCACCTCTCAAGGCTCCCTGCTGCTGGACAT 914
Db 1130 CTTACACCCCTCACCATCGCCAGCAGATCACCTCTCTCAAGGCTCCCTGCTGCTGGACAT 1189
Qy 915 CTTGATCTCTCGGATCTGCACGCGGTACACGCCGAGCAGGACACCATGACCTTCTCGGA 974
Db 1190 CTTGATCTCTCGGATCTGCACGCGGTACACGCCGAGCAGGACACCATGACCTTCTCGGA 1249
Qy 975 CGGGCTGACCTGAAACCGGACCCAGATGCACACGCTGGCTTGGGCCCTTCACCGACCT 1034
Db 1250 CGGGCTGACCTGAAACCGGACCCAGATGCACACGCTGGCTTGGGCCCTTCACCGACCT 1309
Qy 1035 GGTCTTTGCTTCCCAACACAGCTGCTGCCCTGGAGATGGATGATGATGATGATGATGATGAT 1094
Db 1310 GGTCTTTGCTTCCCAACACAGCTGCTGCCCTGGAGATGGATGATGATGATGATGATGATGAT 1369
Qy 1095 GCTCAGCGCCATCTGCTCATCTCGCGAGACCCGACGACCTGGAGCAGCCGCGGACCGGGT 1154
Db 1370 GCTCAGCGCCATCTGCTCATCTCGCGAGACCCGACGACCTGGAGCAGCCGCGGACCGGGT 1429
Qy 1155 GGACATGCTCAGAGAGCCGCTGTGGAGGCGCTAAAGGTCTACGTGGGAAAGCGGAGGCC 1214
Db 1430 GGACATGCTCAGAGAGCCGCTGTGGAGGCGCTAAAGGTCTACGTGGGAAAGCGGAGGCC 1489
Qy 1215 CAGCGCCGCCCATGTTTCCCAAGATGCTAATGAAGATTAATGCTGCTGCTGCTGCTGCTGCT 1274
Db 1490 CAGCGCCGCCCATGTTTCCCAAGATGCTAATGAAGATTAATGCTGCTGCTGCTGCTGCTGCT 1549
Qy 1275 CGCCAAAGGGGCTGAGCGGGGTGATCAGCTGAAGATGGAGATCCCGGCTCCATGCCGCC 1334
Db 1550 CGCCAAAGGGGCTGAGCGGGGTGATCAGCTGAAGATGGAGATCCCGGCTCCATGCCGCC 1609
```

```
QY 1335 TCTCATCCAGGAATGTTGGAGAACTCAGAGGGCTGGACACTCTGACCGGACAGCCGGG 1394
Db 1610 TCTCATCCAGGAATGTTGGAGAACTCAGAGGGCTGGACACTCTGACCGGACAGCCGGG 1669
QY 1395 GGTGGGGGGGGGACGGGGTGGCTTGGCCCGCCCGCCAGGAGCTGTAGCCCGAGCCT 1454
Db 1670 GGTGGGGGGGGGACGGGGTGGCTTGGCCCGCCCGCCAGGAGCTGTAGCCCGAGCCT 1729
QY 1455 CAGCCCCAGCTCCCAACAGAGCAGCCGGCCACCACTCCCGGTGACCGCCAGCCACA 1514
Db 1730 CAGCCCCAGCTCCCAACAGAGCAGCCGGCCACCACTCCCGGTGACCGCCAGCCACA 1789
QY 1515 TGGACACAGCCCTCGCCCTCGCCCGCCGGCTTTCTCTGCTTTCTACCGACCATGTGACC 1574
Db 1790 TGGACACAGCCCTCGCCCTCGCCCGCCGGCTTTCTCTGCTTTCTACCGACCATGTGACC 1849
QY 1575 CCGCACAGCCCTCGCCCGCCAGCTCCCTCCCGGGGAGTACTGGGGACCTTCCCTGGGGGA 1634
Db 1850 CCGCACAGCCCTCGCCCGCCAGCTCCCTCCCGGGGAGTACTGGGGACCTTCCCTGGGGGA 1909
QY 1635 CGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1694
Db 1910 CGGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1969
QY 1695 CTCCACAGCCCTGGGCTGACGCTCAGAGCCGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAG 1754
Db 1970 CTCCACAGCCCTGGGCTGACGCTCAGAGCCGAGGCGAGGAGGAGGAGGAGGAGGAGGAGGAG 2029
QY 1755 CTGGGCTCAGGATGGGTCTGGGGGCTCGTGTTCATCAAGAC 1798
Db 2030 CTGGGCTCAGGATGGGTCTGGGGGCTCGTGTTCATCAAGAC 2073

RESULT 14
HSRAR
LOCUS
DEFINITION Human mRNA for retinoic acid receptor.
ACCESSION X06538
VERSION X06538.1 GI:35873
KEYWORDS DNA binding protein; hormone receptor; receptor; retinoic acid receptor.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 260 to 1920)
AUTHORS Petkovich, M., Brand, N.J., Krust, A. and Chambon, P.
TITLE A human retinoic acid receptor which belongs to the family of nuclear receptors
JOURNAL Nature 330 (6147), 444-450 (1987)
MEDLINE 88063872
PUBMED 2825025
REFERENCE 2 (bases 1 to 420)
AUTHORS Chambon, P.
TITLE Direct Submission
JOURNAL Submitted (22-DEC-1988)
COMMENT cell line-Mcf-7; library-lambda gt10; clone-p63.
FEATURES
    Location/Qualifiers
        1..1920
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            317..1615
            /note="retinoic acid receptor (AA 1 - 432)"
            /codon_start=1
            /protein_id="CAA29787.1"
            /db_xref="GI:35873"
            /db_xref="SWISS-PROT:P10276"
            /translation="MLGSLPPGALFTLQHLVPSVSTPSTPIETQSSSEIIVPS
PPSPPLRIYKPCFVQDKSSGIHYGSACEGKGFFRRSIQKNMYVTCRDKNCII
NKYTRNRCQYLRQKCFVGMGSKESYVNRDNKKKKEVPKPESESYLTLPVEGELIK
VRKAHETFPALQAKYITNNSEQRSLDIDLWDKFSLSLTKCIITKTVFAKQLPG
FTTLTIADQITLLAKACLDLILIRICTRYTPEQDTWTFSGLTLNRTMHNAGFGPLT
```

```
DLVAFANQLPLEMDAETGLLSAICLICDRODLEQPDVRDMLQELLEALKVYVR
KRRSPRHMPFKMLKTDLSRSISAKGAERVITLKMETPGSMPLIOEMLENSEGLDT
LSGPGGGGGGGGLAPPGGSPSLSPSSNRSSPATHSP"
488..685
/note="put. DNA-binding domain"
763..764
/note="gc was cg in [1]"
/citation=[1]
824..1483
/note="put. ligand-binding domain"
1395
/note="g was c in [1]"
/citation=[1]
1733
/note="g was c in [1]"
/citation=[1]
BASE COUNT 386 a 638 c 566 g 330 t
ORIGIN

Query Match 71.9%; Score 1499.8; DB 9; Length 1920;
Best Local Similarity 99.5%; Pred. No. 1.1e-261;
Matches 1504; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 288 CTCCATTGAGACCCAGAGCAGCAGTTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACC 347
Db 403 CACCATTGAGACCCAGAGCAGCAGTTCTGAAGAGATAGTCCCGAGCCCTCCCTCGCCACC 462
QY 348 CCCTACCCCGCATCTACAAGCCTTGTCTGTGTAGGACAAGTCTCAGGCTACCA 407
Db 463 CCCTACCCCGCATCTACAAGCCTTGTCTGTGTAGGACAAGTCTCAGGCTACCA 522
QY 408 CTATGGGCTACGCGCTGTGAGGGCTGCAAGGGCTTCTTCGCGCCGAGCATCCAGAGAA 467
Db 523 CTATGGGCTACGCGCTGTGAGGGCTGCAAGGGCTTCTTCGCGCCGAGCATCCAGAGAA 582
QY 468 CATGTTGTACAGTGTACCCGGGACAAAGATGTCATCAACAAGGTGACCCGGGAACC 527
Db 593 CATGTTGTACAGTGTACCCGGGACAAAGATGTCATCAACAAGGTGACCCGGGAACC 642
QY 528 CTGCGAGTACTGCCGACTGCGAGAGTGTCTTGAAGTGGGATGTCCAGGAGTCTGTGAG 587
Db 643 CTGCGAGTACTGCCGACTGCGAGAGTGTCTTGAAGTGGGATGTCCAGGAGTCTGTGAG 702
QY 588 AAACGACCGGAACAGAAAGAGAGGAGTGCCTCAAGCCGAGTGTCTGTGAGAGCTACAC 647
Db 703 AAACGACCGGAACAGAAAGAGAGGAGTGCCTCAAGCCGAGTGTCTGTGAGAGCTACAC 762
QY 648 GCTGACCGCGGAGTGGGGAGCTCATTTGAGAAGGTGCGCAAGGCGCACAGGAAACCTT 707
Db 763 GCTGACCGCGGAGTGGGGAGCTCATTTGAGAAGGTGCGCAAGGCGCACAGGAAACCTT 822
QY 708 CCTGCGCTCTGCCAGTGGGCAATATACACTACGAAACAGCTCAGAACACAGTGTCTC 767
Db 823 CCTGCGCTCTGCCAGTGGGCAATATACACTACGAAACAGCTCAGAACACAGTGTCTC 882
QY 768 TCTGACATTCACCTCTCGGACAGTTCAGTGAAGTCTCCACCAAGTGCATCATTAAGAC 827
Db 882 TCTGACATTCACCTCTCGGACAGTTCAGTGAAGTCTCCACCAAGTGCATCATTAAGAC 942
QY 828 TGTGAGTTCGCCAGAGCTGCCCGGCTTCAACCCCTTCAACCCCTTCAACCCCTTCAACCCCT 887
Db 943 TGTGAGTTCGCCAGAGCTGCCCGGCTTCAACCCCTTCAACCCCTTCAACCCCTTCAACCCCT 1002
QY 888 CCTGCTCAAGGCTGCTGCTGGGACATCTGTATCTGGGATCTGACCGCGGTACAGCC 947
Db 1003 CCTGCTCAAGGCTGCTGCTGGGACATCTGTATCTGGGATCTGACCGCGGTACAGCC 1062
QY 948 CGAGCAGACACCATGACCTTCTCGGAGGGCTGACCTGAACCGGACCCAGATGACAA 1007
Db 1063 CGAGCAGACACCATGACCTTCTCGGAGGGCTGACCTGAACCGGACCCAGATGACAA 1122
QY 1008 CGCTGGCTTCGCGCCCTTCAACGAGCTGTCTTTTGCCTTCGCGCAACAGCTGTGCCCCCT 1067
```

Db	1123	CGTGGCTTGGCCCTCACCAGACCTGGTCTTTGGCTTCGCCAACCAACAGCTGCTGCCCT	1182
Qy	1068	GGAGATGATGATGCGGAGAGGGGCTGCTCAGCCGCCATCTCCCTCATCTCGCGAGACGG	1127
Db	1183	GGAGATGATGATGCGGAGAGGGGCTGCTCAGCCGCCATCTCCCTCATCTCGCGAGACGG	1242
Qy	1128	CCAGACCTGGAGCAGCGGAGCGGCTGGACATGCTGCAGGAGCGGCTGCTGGAGCGCT	1187
Db	1243	CCAGACCTGGAGCAGCGGAGCGGCTGGACATGCTGCAGGAGCGGCTGCTGGAGCGCT	1302
Qy	1188	AAAGTCTACGTGCGGAAGCGGAGCGGCTGGACATGCTGCAGGAGCGGCTGCTGGAGCGCT	1247
Db	1303	AAAGTCTACGTGCGGAAGCGGAGCGGCTGGACATGCTGCAGGAGCGGCTGCTGGAGCGCT	1362
Qy	1248	GAAGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGGCTGAGCGGGTGATCAGCGTAA	1307
Db	1363	GAAGATTACTGACCTGCGAAGCATCAGCGCCCAAGGGGGCTGAGCGGGTGATCAGCGTAA	1422
Qy	1308	GATGAGATCCGGGCTCCATCGCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGG	1367
Db	1423	GATGAGATCCGGGCTCCATCGCGCTCTCATCCAGGAAATGTTGGAGAACTCAGAGG	1482
Qy	1368	CCTGACACTCTGACGCGACAGCGGGGGTGGGGGGGACGGGGTGGCTTGCCTCC	1427
Db	1483	CCTGACACTCTGACGCGACAGCGGGGGTGGGGGGGACGGGGTGGCTTGCCTCC	1542
Qy	1428	CCGCGCAGGAGCTGTAGCGCCAGCGCTCAGCGCCAGCTCCACAGAGCAGCGCGGCCAC	1487
Db	1543	CCGCGCAGGAGCTGTAGCGCCAGCGCTCAGCGCCAGCTCCACAGAGCAGCGCGGCCAC	1602
Qy	1488	CCACTCCCGTGAACCGCCACGCCACATGAGACAGCCCTCGCCCTCCCGCGGGCTTTT	1547
Db	1603	CCACTCCCGTGAACCGCCACGCCACATGAGACAGCCCTCGCCCTCCCGCGGGCTTTT	1662
Qy	1548	CTCTGCTTTTACGACCATGTGACCGCGACCGCTGCGCCACCTGCGCTCCCGG	1607
Db	1663	CTCTGCTTTTACGACCATGTGACCGCGACCGCTGCGCCACCTGCGCTCCCGG	1722
Qy	1608	CGAGTACTGGGACCTTCCCTGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1667
Db	1723	CGAGTACTGGGACCTTCCCTGGGGACGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	1782
Qy	1668	AGGCTTGGGCTTACGTGAGCTGCTGCTCCACAGCTTGGCTGAGCTGAGGCGCGAG	1727
Db	1783	AGGCTTGGGCTTACGTGAGCTGCTGCTCCACAGCTTGGCTGAGCTGAGGCGCGAG	1842
Qy	1728	GCAGAACTGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1787
Db	1843	GCAGAACTGAGTGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	1902
Qy	1788	TTTCAACAGAC	1798
Db	1903	TTTCAACAGAC	1913
RESULT	15		
LOCUS	HSU41743		
DEFINITION	Human nucleophosmin-retinoic acid receptor alpha fusion protein		
ACCESSION	U41743		
VERSION	U41743.1		
KEYWORDS	GI:1314309		
SOURCE	Homo sapiens		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 1944)		
TITLE	Redner, R.L., Rush, E.A., Faas, S., Rudert, W.A. and Corey, S.J.		
JOURNAL	The t(5;17) variant of acute promyelocytic leukemia expresses a nucleophosmin-retinoic acid receptor fusion		
MEDLINE	Blood 87 (3), 882-886 (1996)		

8562957	2 (bases 1 to 1944)	Redner, R.L.	Direct Submission	Submitted (30-NOV-1995)	Robert L. Redner, Medicine, University of Pittsburgh, E1058 BST, 211 Lothrop Street, Pittsburgh, PA 15213, USA
FEATURES	Location/Qualifiers				
source	1..1944				
	/organism="Homo sapiens"				
	/db_xref="taxon:9606"				
	/cell_type="acute promyelocytic leukemia cells with t(5;17)(q32;q11) translocation"				
	/tissue_type="bone marrow"				
	84..1646				
	/codon_start=1				
	/product="nucleophosmin-retinoic acid receptor alpha fusion protein NPM-RAR short form"				
	/protein_id="AAB00113.1"				
	/db_xref="GI:1314310"				
	/translation="MEDSDMDMSPLRPONYLFGCELKADKDYHFKVDNENEHQLSLRTVSGAGAKDELHIVAEAMNYEGSPIKVTLATKMSVQPTVSLGFEITPPVLRKKGSGVHISGQHLVAIETSSSEIEVPSPPLPRIYKPCFQCDKSSYHGVSAECGKGFRRSIQNMVYTCRDKNCIIINVRKQYCRQKQKFEVGMSSSEQRVDRNKKKVPKPECSESYLTPEVGLIEBKVKAHOETPALCOLGKYKTYNNSEQRVSLDIDLNDKFSLSKCIITVEFAKQLPGFTTLTADQITLLKARCLDILLIRICTRYPEQDTMTFSDGLTNRTQMNHAGGLTDLVAFANLLPLEMDDADETGLLSAICLIGCDRLQPDQPDVDMLOEPLLEALKVYVRRKRRSPHMFPMKMDITDLSISAAGAERVITLKEIPGSMPLPIQEMLENSGLDLSQPGGGGRDGGGLAPPPGSCPSLSPSSNRSPATHSP"				
CDS	427 a 598 c 557 g 362 t				
BASE COUNT	427 a 598 c 557 g 362 t				
ORIGIN					
	Query Match 71.9%; Score 1499.4; DB 9; Length 1944;				
	Best Local Similarity 99.6%; Pred. No. 1.3e-261;				
	Matches 1503; Conservative 0; Mismatches 6; Indels 0; Gaps 0;				
Qy	290	CCATTGAGACCCAGAGCAGCAGTCTGAGACAGATAGTGCCCGCCCTCCCTCGCCACCCC	349		
Db	436	CCATTGAGACCCAGAGCAGCAGTCTGAGACAGATAGTGCCCGCCCTCCCTCGCCACCCC	495		
Qy	350	CTCTACCCCGCATCTACAAGCCTTGCTTCTGTCTAGGACAAGTCTCTCAGGCTACCACT	409		
Db	496	CTCTACCCCGCATCTACAAGCCTTGCTTCTGTCTAGGACAAGTCTCTCAGGCTACCACT	555		
Qy	410	ATGGGGTTCAGCGCTGTGAGGGCTGCAAGGGCTTCTTCCGCGCAGCATCCAGAGAACA	469		
Db	556	ATGGGGTTCAGCGCTGTGAGGGCTGCAAGGGCTTCTTCCGCGCAGCATCCAGAGAACA	615		
Qy	470	TGTTGTACAGTGTACCGGGACAGAAGTGCATCATCAAGAGTGCACCGGACCCCT	529		
Db	616	TGTTGTACAGTGTACCGGGACAGAAGTGCATCATCAAGAGTGCACCGGACCCCT	675		
Qy	530	GCCAGTACTCCGACTGCAGAGTGTCTTGAAGTGGCATGTCCCAAGAGTCTGTGAGAA	589		
Db	676	GCCAGTACTCCGACTGCAGAGTGTCTTGAAGTGGCATGTCCCAAGAGTCTGTGAGAA	735		
Qy	590	ACGACCGAAAACAGAAGAAGAGGAGTGCCCAAGCCCGAGTGTCTGTGAGAGTACACGC	649		
Db	736	ACGACCGAAAACAGAAGAAGAGGAGTGCCCAAGCCCGAGTGTCTGTGAGAGTACACGC	795		
Qy	650	TGACCGCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGCAAGCCACACAGGAAACCTTCC	709		
Db	796	TGACCGCGGAGGTGGGGAGCTCATTTGAGAAGGTGCGCAAGCCACACAGGAAACCTTCC	855		
Qy	710	CTGCCCTCTCCAGCTGGGCAAAATACACTAGCAACACAGCTCAGAACACAGTGTCTCTC	769		
Db	856	CTGCCCTCTCCAGCTGGGCAAAATACACTAGCAACACAGCTCAGAACACAGTGTCTCTC	915		
Qy	770	TGCACATTGACCTCTGGGCAAGTTCAGTGAACCTCCACCAAGTGCATCATTAAGACTG	829		
Db	916	TGCACATTGACCTCTGGGCAAGTTCAGTGAACCTCCACCAAGTGCATCATTAAGACTG	975		







GenCore version 5.1.4.p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2003, 05:30:18 ; Search time 62 seconds  
(without alignments)  
1518.767 Million cell updates/sec

Title: US-09-691-220-2

Perfect score: 2437

Sequence: 1 MYSEVEGGTPNPFVLVDF.....SCSPSLSPSSNRSSPATHSP 457

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL 21.\*

- 1: sp\_archaea.\*
- 2: sp\_bacteria.\*
- 3: sp\_fungi.\*
- 4: sp\_human.\*
- 5: sp\_invertebrate.\*
- 6: sp\_mammal.\*
- 7: sp\_mhc.\*
- 8: sp\_organelle.\*
- 9: sp\_phage.\*
- 10: sp\_plant.\*
- 11: sp\_rodent.\*
- 12: sp\_virus.\*
- 13: sp\_vertebrate.\*
- 14: sp\_unclassified.\*
- 15: sp\_virus.\*
- 16: sp\_bacteriap.\*
- 17: sp\_archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2349	96.4	459	11 Q9QWJ1	Q9QWJ1 rattus norv
2	2112	86.7	797	4 Q15156	Q15156 homo sapien
3	2092	85.8	462	11 P97513	P97513 mus spretus
4	2072.5	85.0	430	11 Q8VHB8	Q8VHB8 mesocricetu
5	2026.5	83.2	448	13 Q92019	Q92019 xenopus lae
6	2021	82.9	457	13 Q90272	Q90272 brachydanio
7	2008.5	82.4	444	13 Q91391	Q91391 brachydanio
8	1998.5	82.0	444	13 Q90271	Q90271 brachydanio
9	1770	72.6	448	11 Q8VHB7	Q8VHB7 mesocricetu
10	1760	72.2	448	13 Q910C7	Q910C7 coturnix co
11	1651	67.7	458	11 Q91YX2	Q91YX2 mus musculu
12	1651	67.7	458	11 Q91VK5	Q91VK5 mus musculu
13	1646.5	67.6	443	11 Q8VHB6	Q8VHB6 mesocricetu
14	1641	67.3	441	13 Q918T3	Q918T3 ambystoma m
15	1628.5	66.8	446	13 Q918T2	Q918T2 ambystoma m
16	1620	66.5	307	11 Q9QWT3	Q9QWT3 rattus norv

17	1613	66.2	426	13 Q90968	Q90968 gallus gall
18	1595.5	65.5	381	4 Q9UJ38	Q9UJ38 homo sapien
19	1474.5	60.5	367	13 Q90273	Q90273 brachydanio
20	1387	56.9	302	11 Q9QWT4	Q9QWT4 rattus norv
21	1358	55.7	285	13 Q90Y68	Q90Y68 paralichthy
22	1335	54.8	285	13 Q90Y67	Q90Y67 paralichthy
23	1290	52.9	279	13 Q9DQJ6	Q9DQJ6 salmo salar
24	1287.5	52.8	286	13 Q90Y70	Q90Y70 paralichthy
25	1234	50.6	287	13 Q90Y69	Q90Y69 paralichthy
26	1173	48.1	434	5 P91780	P91780 polyandroca
27	1151	47.2	291	11 Q54746	Q54746 rattus norv
28	988	40.5	347	5 P91779	P91779 polyandroca
29	950	39.0	224	6 P79300	P79300 sus scrofa
30	683	28.0	130	6 P79298	P79298 sus scrofa
31	652	26.8	410	4 Q96H73	Q96H73 homo sapien
32	627	25.7	141	6 P79297	P79297 sus scrofa
33	625	25.6	387	13 Q90Y21	Q90Y21 salmo salar
34	625	25.6	555	15 Q96594	Q96594 avian eryth
35	625	25.6	582	15 Q96593	Q96593 avian eryth
36	623.5	25.6	396	13 Q90Y22	Q90Y22 salmo salar
37	623.5	25.6	455	15 Q85511	Q85511 avian eryth
38	617	25.3	555	15 Q36200	Q36200 avian eryth
39	612.5	25.1	962	15 Q84895	Q84895 avian eryth
40	562.5	23.1	321	13 Q90Y30	Q90Y30 fugu rubrip
41	553.5	22.7	445	11 Q91X41	Q91X41 mus musculu
42	537	22.0	412	13 Q91840	Q91840 xenopus lae
43	536.5	22.0	545	5 Q9U4L1	Q9U4L1 aedes aegyp
44	534	22.0	149	13 Q918T4	Q918T4 ambystoma m
45	532	21.8	329	13 Q90Y19	Q90Y19 scophthalmu

#### ALIGNMENTS

#### RESULT 1

Q9QWJ1 PRELIMINARY; PRT; 459 AA.

AC Q9QWJ1: Q9QWJ1: 459 AA.  
DT 01-MAY-2000 (Tremblrel. 13, Created)  
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)  
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
DE Retinoic acid receptor alpha 2 isoform.  
OS RAR.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-TESTIS;  
RX MEDLINE=96301334; PubMed=8722633;  
RA Akmal K.M., Dufour J.M., Kim K.H.;  
RT "Region-specific localization of retinoic acid receptor-alpha expression in the rat epididymis";  
RL Biol. Reprod. 54:1111-1119(1996).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-TESTIS;  
RX MEDLINE=97167633; PubMed=9116160;  
RA Akmal K.M., Dufour J.M., Kim K.H.;  
RT "Retinoic acid receptor alpha gene expression in the rat testis: potential role during the prophase of meiosis and in the transition from round to elongating spermatids";  
RL Biol. Reprod. 56:549-556(1997).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE-DAWLEY; TISSUE-TESTIS;  
RX MEDLINE=98151023; PubMed=9492059;  
RA Akmal K.M., Dufour J.M., Vo M., Higginson S., Kim K.H.;  
RT "Ligand-dependent regulation of retinoic acid receptor alpha in rat testis: in vivo response to depletion and repletion of vitamin A";  
RL Endocrinology 139:1239-1248(1998).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).

CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
DR EMBL; U15211; AAC23439.1; -.  
DR HSSP; P10826; 1HRA.  
DR InterPro; IPR000536; Hormone\_rec\_lig.  
DR InterPro; IPR001723; Stdhrmn\_receptor.  
DR InterPro; IPR001628; Znf\_C4steroid.  
DR Pfam; PF00104; hormone\_rec; 1.  
DR Pfam; PF00105; zf-C4; 1.  
DR PRINTS; PR00398; STRDHORMONER.  
DR PRINTS; PR00047; STROIDFINGER.  
DR ProDom; PD000035; Znf\_C4steroid; 1.  
DR SMART; SM00430; HOLI; 1.  
DR SMART; SM00399; Znf\_C4; 1.  
DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;  
KW Zinc-finger.  
SQ SEQUENCE 459 AA; 50928 MW; F03368934C1E3C86 CRC64;  
  
Query Match 96.4%; Score 2349; DB 11; Length 459;  
Best Local Similarity 96.9%; Pred. No. 2.7e-195;  
Matches 445; Conservative 4; Mismatches 8; Indels 2; Gaps 1;  
  
QY 1 MYSEVEVG--PTPNFVVDYFNQNRACLLPEKGLPAGPGYSTPLRTPLWNGSNHSIET 58  
DB 1 MYSEVEVGGLTPAPNPFVVDYFNQNRACLLPEKGLPAGPGYSTPLRTPLWNGSNHSIET 60  
  
QY 59 QSSSESEIVSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRRSIQKNVYT 118  
DB 61 QSSSESEIVSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRRSIQKNVYT 120  
  
QY 119 CHRDKNCIINKVTRNFCQYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPCSESYTLTPE 178  
DB 121 CHRDKNCIINKVTRNFCQYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPCSESYTLTPE 180  
  
QY 179 VGLIEKVRKAHOETFPALCQKGYTNNSEQRVSLDIDLWDKFSLSKCIKTVEFA 238  
DB 181 VGLIEKVRKAHOETFPALCQKGYTNNSEQRVSLDIDLWDKFSLSKCIKTVEFA 240  
  
QY 239 KQLPGFTTTLTADQITLLKAAACLDLILTRICTRYTPEDQTMFTSDGLTLNRTQMHNAGF 298  
DB 241 KQLPGFTTTLTADQITLLKAAACLDLILTRICTRYTPEDQTMFTSDGLTLNRTQMHNAGF 300  
  
QY 299 PLTDLVFAFANQLPLEMDDAETGLLSAICLICGDRQDLQPDQVDMQLQEPLEALKVYV 358  
DB 301 PLTDLVFAFANQLPLEMDDAETGLLSAICLICGDRQDLQPDQVDMQLQEPLEALKVYV 360  
  
QY 359 RRRRSPRPHFPKMLMKITDLRSISAKGAERVITLKMIEPGSMPLLIQEMLENSEGLDTL 418  
DB 361 RRRRSPRPHFPKMLMKITDLRSISAKGAERVITLKMIEPGSMPLLIQEMLENSEGLDTL 420  
  
QY 419 SGQPGGGGRDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
DB 421 SGQPGGGGRDGGGLPPPGSCPSLSPSSNRSSPATHSP 459  
  
RESULT 2  
Q15156 PRELIMINARY; PRT; 797 AA.  
AC Q15156;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE PML-RAR protein.  
GN PML-RAR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=91347368; PubMed=16523368;  
RA Kakizuka A., Miller W.H. Jr., Umesono K., Warrell R.P. Jr.,  
RA Frankel S.R., Murty V.V., Dmitrovsky E., Evans R.M.;

\*Chromosomal translocation t(15;17) in human acute promyelocytic leukemia fuses RAR alpha with a novel putative transcription factor, PML\*;  
Cell 66:663-674(1991).  
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
CC -!- SIMILARITY: CONTAINS 1 RING-TYPE ZINC FINGER.  
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
DR EMBL; M73779; AAA60126.1; -.  
DR HSSP; P29590; 1BOR.  
DR InterPro; IPR000536; Hormone\_rec\_lig.  
DR InterPro; IPR001723; Stdhrmn\_receptor.  
DR InterPro; IPR000315; Znf\_Box.  
DR InterPro; IPR001628; Znf\_C4steroid.  
DR InterPro; IPR001841; Znf\_ring.  
DR Pfam; PF00104; hormone\_rec; 1.  
DR Pfam; PF00643; zf-B\_box; 2.  
DR Pfam; PF00097; zf-C3HC4; 1.  
DR Pfam; PF00105; zf-C4; 1.  
DR PRINTS; PR00398; STRDHORMONER.  
DR PRINTS; PR00047; STROIDFINGER.  
DR ProDom; PD000035; Znf\_C4steroid; 1.  
DR SMART; SM00336; BBOX; 1.  
DR SMART; SM00430; HOLI; 1.  
DR SMART; SM00184; RING; 1.  
DR SMART; SM00399; Znf\_C4; 1.  
DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
DR PROSITE; PS00518; ZF\_RING\_1; 1.  
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;  
KW Zinc-finger.  
SQ SEQUENCE 797 AA; 89281 MW; 8C16B6E85CBBDD214 CRC64;  
  
Query Match 86.7%; Score 2112; DB 4; Length 797;  
Best Local Similarity 99.3%; Pred. No. 2e-174;  
Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
  
QY 55 SIETQSSSEIVSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRRSIQKN 114  
DB 395 AIETQSSSEIVSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRRSIQKN 454  
  
QY 115 MYTCHRDKNCIINKVTRNFCQYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPCSESYT 174  
DB 455 MYTCHRDKNCIINKVTRNFCQYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPCSESYT 514  
  
QY 175 LTPVEGLIEKVRKAHOETFPALCQKGYTNNSEQRVSLDIDLWDKFSLSKCIKT 234  
DB 515 LTPVEGLIEKVRKAHOETFPALCQKGYTNNSEQRVSLDIDLWDKFSLSKCIKT 574  
  
QY 235 VEFKQLPGFTTTLTADQITLLKAAACLDLILTRICTRYTPEDQTMFTSDGLTLNRTQMHN 294  
DB 575 VEFKQLPGFTTTLTADQITLLKAAACLDLILTRICTRYTPEDQTMFTSDGLTLNRTQMHN 634  
  
QY 295 AGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLICGDRQDLQPDQVDMQLQEPLEAL 354  
DB 635 AGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLICGDRQDLQPDQVDMQLQEPLEAL 694  
  
QY 355 KVYVRRRSPRPHFPKMLMKITDLRSISAKGAERVITLKMIEPGSMPLLIQEMLENSE 414  
DB 695 KVYVRRRSPRPHFPKMLMKITDLRSISAKGAERVITLKMIEPGSMPLLIQEMLENSE 754  
  
QY 415 LDTLSGGGGGRDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
DB 755 LDTLSGGGGGRDGGGLPPPGSCPSLSPSSNRSSPATHSP 797  
  
RESULT 3  
P97513 PRELIMINARY; PRT; 462 AA.  
ID P97513;  
AC P97513;  
DT 01-MAY-1997 (TREMBLrel. 03, Created)  
DT 01-MAY-1997 (TREMBLrel. 03, Last sequence update)  
DE 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Retinoic acid receptor-alpha.  
GN RARA.



DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Retinoic acid receptor.  
 GN RARA.  
 OS Xenopus laevis (African clawed frog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;  
 OC Xenopodinae; Xenopus.  
 OX NCBI\_TaxID=8355;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93144180; PubMed=1336976;  
 RA Sharpe C.R.;  
 RT "Two isoforms of retinoic acid receptor alpha expressed during Xenopus  
 RT development respond to retinoic acid."  
 RL Mech. Dev. 39:81-93(1992).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Sharpe C.R.;  
 RL Submitted (MAY-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 DR EMBL; X87363; CAA80791.1; -.  
 DR HSSP; P10826; IHRA.  
 DR InterPro; IPR005336; Hormone\_rec\_lig.  
 DR InterPro; IPR001723; Stdhrmn\_receptor.  
 DR Pfam; PF00104; hormone\_rec; 1.  
 DR Pfam; PF00105; zf-C4; 1.  
 DR PRINTS; PR00398; STDRHORMONER.  
 DR PRODOM; PD000035; Znf\_C4steroid; 1.  
 DR SMART; SM00430; HOL1; 1.  
 DR SMART; SM00399; ZNF\_C4; 1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR DNA-binding; Nuclear protein; Receptor; Transcription regulation;  
 KW Zinc-finger.  
 SQ SEQUENCE 448 AA; 50456 MW; 6CF94B31BC57054B CRC64;  
  
 Query Match 83.2%; Score 2026.5; DB 13; Length 448;  
 Best Local Similarity 84.9%; Pred. No. 2.4e-167;  
 Matches 388; Conservative 22; Mismatches 38; Indels 9; Gaps 3;  
 QY 1 MYESVEVGGTPNPFLVDFYNONRACLLPEKGLPAPGPYSTPLRPLWNGSNHSIETQS 60  
 DB 1 MYENVDV---SPHYHMDVSYNRQCLWPEKRI---NPYGTPLGTQHWSSSNHSIETQS 54  
 QY 61 SSSEIVSPSPPPPLPRIYKPCFVCQDKSSGYHYGVSACEGCKGFFRRSIQKNMVTCH 120  
 DB 55 TSSEIVSPSPPPPLPRIYKPCFVCQDKSSGYHYGVSACEGCKGFFRRSIQKNMVTCH 114  
 QY 121 RDKNCLINKVTRNRCQYCRLOKCFEYGVMSKESYVRNDRNKKKKEVPKPECSSEYTLTPEVG 180  
 DB 115 RDKNCLINKVTRNRCQYCRLOKCFEYGVMSKESYVRNDRNKKKKEVPKPECSSEYTLTPEVG 174  
 QY 181 ELIEKVRKAHOETFPALCOLGKYTTNNSRQYSLDLDLWDFSELSTKCIITKVEFAKQ 240  
 DB 175 DLIEKVRKAHOETFPALCOLGKYTTNNSRQYSLDLDLWDFSELSTKCIITKVEFAKQ 234  
 QY 241 LPGFVTLTIADQITLLKAACLDLILIRICTRYTPPEQDTMTFSDGLTLNRTQMAGFPL 300  
 DB 235 LPGFVTLTIADQITLLKAACLDLILIRICTRYTPPEQDTMTFSDGLTLNRTQMAGFPL 294  
 QY 301 TLVFAFANQLPLEMDDAETGLLSAICLICGDRQDLQPDQVDMLEALQVYVRK 360  
 DB 295 TLVFAFANQLPLEMDDAETGLLSAICLICGDRQDLQPDQVDMLEALQVYVRN 354  
 QY 361 RPSRPHMFKMLKTDLRISAKGAERVITLTKMEIPGSMPLIQEMLENSEGLDTLGS 420  
 DB 355 RRPKPHMFKMLKTDLRISAKGAERVITLTKMEIPGSMPLIQEMLENSEGLDTLGS 413  
 QY 421 QPGGGRDGGGLPPPGSCPSLSPSSNRSSPATHSP 457

Db 414 --GGASSQTPVTPVAPGSCSPSLSPSTHSSPSSSP 448  
 RESULT 6  
 Q90272  
 ID Q90272 PRELIMINARY; PRT; 457 AA.  
 AC Q90272;  
 DT 01-NOV-1996 (Tremblrel. 01, Created)  
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
 DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)  
 DE Retinoic acid receptor alpha-2.B.  
 GN RARA2B OR RAR-ALPHA-2.B.  
 OS Brachydanio rerio (zebrafish) (Zebra danio).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Osteichthyes; Cypriniformes;  
 OC Cyprinidae; Danio.  
 OX NCBI\_TaxID=7955;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=EMBRYO;  
 RA Stachel S.E.; Kushner P.;  
 RT "The molecular characterization of three zebrafish retinoic acid  
 RT receptor genes suggests the retinoic acid pathway functions in  
 RT embryonic hindbrain development."  
 RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=EMBRYO;  
 RA Stachel S.E.;  
 RT "Retinoic acid reprograms RAR expression and shows anteroposterior  
 RT specification occurs in the pre-gastrula zebrafish."  
 RL Submitted (DEC-1992) to the EMBL/GenBank/DBJ databases.  
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.  
 DR EMBL; L03399; AAA50050.1; -.  
 DR HSP; P10826; IHRA.  
 DR ZFIN; ZDB-GENE-980526-72; rara2b.  
 DR InterPro; IPR000536; Hormone\_rec\_lig.  
 DR InterPro; IPR001723; Stdhrmn\_receptor.  
 DR InterPro; IPR001628; Znf\_C4steroid.  
 DR Pfam; PF00104; hormone\_rec; 1.  
 DR Pfam; PF00105; zf-C4; 1.  
 DR PRINTS; PR00398; STDRHORMONER.  
 DR PRODOM; PD000035; Znf\_C4steroid; 1.  
 DR SMART; SM00430; HOL1; 1.  
 DR SMART; SM00399; ZNF\_C4; 1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
 DR DNA-binding; Nuclear protein; Receptor; Transcription regulation;  
 KW Zinc-finger.  
 SQ SEQUENCE 457 AA; 50481 MW; DF51E08BD6F0A625 CRC64;  
  
 Query Match 82.9%; Score 2021; DB 13; Length 457;  
 Best Local Similarity 82.4%; Pred. No. 7.4e-167;  
 Matches 378; Conservative 37; Mismatches 40; Indels 4; Gaps 2;  
 QY 1 MYESVEVGG--PTNPFLVDFYNONRACLLPEKGLPAPGPYSTPLRPLWNGSNHSIET 58  
 DB 1 MYESVDVVGTLTPSPNPFLSDYHONRGCLIPDKGLVSGA--ARGFRNPHSGSNHSVET 58  
 QY 59 QSSSEIVSPSPPPPLPRIYKPCFVCQDKSSGYHYGVSACEGCKGFFRRSIQKNMVT 118  
 DB 59 QSTSEIEIVSPSPPPPPRVRKPCFVCQDKSSGYHYGVSACEGCKGFFRRTIQKNMVT 118  
 QY 119 CHRDKNCIIINKVTRNRCQYCRLOKCFEYGVMSKESYVRNDRNKKKKEVPKPECSSEYTLTPE 178  
 DB 119 CHREKSCIIINKVTRNRCQYCRLOKCFEYGVMSKESYVRNDRNKKKKEVPKPECSSEYTLTPE 178  
 QY 179 VGLIEKVRKAHOETFPALCOLGKYTTNNSRQYSLDLDLWDFSELSTKCIITKVEFA 238  
 DB 179 TEXMIEQVRKAHOETFPALCOLGKYTTNNSRQYSLDLDLWDFSELSTKCIITKVEFA 238



Query Match 82.0%; Score 1998.5; DB 13; Length 444;  
 Best Local Similarity 82.5%; Pred. No. 6.4e-165;  
 Matches 377; Conservative 36; Mismatches 31; Indels 13; Gaps 3;

```

QY 1 MYESVEVGGPTNPFLVVDYFNQNRACLLPEKGLPAPGPYSTPLRTPLWNGSNHSHIETQS 60
Db 1 MIESVDV-----NPLMDYDYNQSGCLIPDK---MHPHFSSIRHQHWSGNSHSHIETQS 52
QY 61 SSSEIEIVSPSPPPPLPRIPYKPCFVQCDKSSGHHYGVSAEGCGKGFRRRSIQKNMYYTCH 120
Db 53 TSSEIEIVSPSPPPPPRIYKPCFVQCDKSSGHHYGVSAEGCGKGFRRRSIQKNMYYTCH 112
QY 121 RDKNCIINKVTRNCPQCYCRLOKCFEYVMSKESVNRDRNKKKEVPKPCSESYTLTPEVG 180
Db 113 REKNCIINKVTRNCPQCYCRLOKCFEYVMSKESVNRDRNKKKEVPKPCSESYTLTPEVG 172
QY 181 ELIEKVRKAHQETFPALCOLGKYTTNNSSEORVSLDLDWKFSELSKCIKTIKVEFAKQ 240
Db 173 QMIDRVKKAHQETFPALCOLGKYTTNNSSEORVSLDLDWKFSELSKCIKTIKVEFAKQ 232
QY 241 LPGFTTLTIADQITLLKAACLDIILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGL 300
Db 233 LPGFTTLTIADQITLLKAACLDIILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGL 292
QY 301 TDLVFAFANQLLPLEMDAETGLLSAICLICGDRQDLEQPDVDMLOEPLLEALKVYVRK 360
Db 293 TDLVFAFANQLLPLEMDAETGLLSAICLICGDRQDLEQPDVDMLOEPLLEALKVYVRK 352
QY 361 RRPSPHPKMLKIYDLRSISAKGAERVITLKMIEIPGSMPLPQIOMLENSGLDITLUG 420
Db 353 RRPSPHPKMLKIYDLRSISAKGAERVITLKMIEIPGSMPLPQIOMLENSGLDITLUG 412
QY 421 QPGQGGGRRGGGLPPPPGSCPSLSPSSNRSSPATHSP 457
Db 413 AQGRAS-----ATPGSCPSLSPNQAQSPPTQSP 444

RESULT 9
Q8VHB7
ID Q8VHB7 PRELIMINARY; PRT; 448 AA.
AC Q8VHB7
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Retinoic acid receptor beta-2.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Sharma R.P., McGraw R.A., Dugyala R.R.;
RT "Cloning and characterization of hamster fetal retinoic acid receptor
RT isoforms."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY046944; AAL02361.1;
DR InterPro: IPR000536; Hormone_rec_1lg.
DR InterPro: IPR001723; Stdrhma_receptor.
DR InterPro: IPR001628; Znf_C4steroid.
DR Pfam: PF00104; hormone_rec; 1.
DR Pfam: PF00105; zf-C4; 1.
DR PRINTS: PR00398; STDRHMONER.
DR PRINTS: PR00047; STROIDFINGER.
DR ProDom: PD000035; Znf_C4steroid; 1.
DR SMART: SM00430; HOLI; 1.
DR SMART: SM00399; Znf_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW Receptor.
SQ SEQUENCE 448 AA; 50353 MW; FC4BBEF9F7A69684 CRC64;

Query Match 72.6%; Score 1770; DB 11; Length 448;
Best Local Similarity 74.6%; Pred. No. 4.2e-145;
Matches 343; Conservative 41; Mismatches 58; Indels 18; Gaps 5;

```

Matches 343; Conservative 41; Mismatches 58; Indels 18; Gaps 5;

```

QY 1 MYESVEVGGPTNPFLVVDYFNQNRACLLPEKGLPAPGPYSTPLRTPLWNGSN--HSIE 57
Db 1 MFDGMDVLSVSPGQ--ILDFTASPSQMLQEKALKACLSGFTQAE---WOHRHTAQISIE 55
QY 58 TOSSEIEIVSPSPPPPLPRIPYKPCFVQCDKSSGHHYGVSAEGCGKGFRRRSIQKNMYY 117
Db 56 TOSTSEIEIVSPSPPPPLPRIPYKPCFVQCDKSSGHHYGVSAEGCGKGFRRRSIQKNMYY 115
QY 118 TCHRNKCIINKVTRNCPQCYCRLOKCFEYVMSKESVNRDRNKKKEVPKPCSESYTLTP 177
Db 116 TCHRNKCIINKVTRNCPQCYCRLOKCFEYVMSKESVNRDRNKKKEVPKPCSESYEMTA 175
QY 178 EVGELIEKVRKAHQETFPALCOLGKYTTNNSSEORVSLDLDWKFSELSKCIKTIKVEF 237
Db 176 EDDTEKIRKAHQETFPALCOLGKYTTNNSADHRVLDGLWDFSELSKCIKTIKVEF 235
QY 238 AKQLPGFTTLTIADQITLLKAACLDIILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAG 297
Db 236 AKRLPGFTSLTIADQITLLKAACLDIILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAG 295
QY 298 GPLTDLVFAFANQLLPLEMDAETGLLSAICLICGDRQDLEQPDVDMLOEPLLEALKVY 357
Db 296 GPLTDLVFAFANQLLPLEMDAETGLLSAICLICGDRQDLEQPDVDMLOEPLLEALKVY 355
QY 358 VYKRRSPHPKMLKIYDLRSISAKGAERVITLKMIEIPGSMPLPQIOMLENSGLDIT 417
Db 356 IKRRSPHPKMLKIYDLRSISAKGAERVITLKMIEIPGSMPLPQIOMLENSGLDIT 415
QY 418 LSGQGGGRRGGGLPPPPGSCPSLSPSSNRSSPATHSP 457
Db 416 LTPSSSGNTAEH-----SPSVSPSSVNSGVQSOP 445

RESULT 10
Q910C7
ID Q910C7 PRELIMINARY; PRT; 448 AA.
AC Q910C7
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Retinoic acid receptor beta-2.
GN RAR-BETA-2.
OS Coturnix coturnix (common quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae;
OC Coturnix.
OX NCBI_TaxID=9091;
RN [1]
RP SEQUENCE FROM N.A.
RA Fu Z., Kubo T., Sugahara K., Noguchi T., Kato H.;
RT "Cloning of Complementary Deoxyribonucleic Acids Encoding Quail
RT Coturnix coturnix japonica) Retinoic Acid Receptor beta Isoforms and
RT Changes in their Gene Expression During Gonadotropic Growth."
RL Biol. Reprod. 64:231-241(2001).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL: AF110730; AAD23398.1;
DR InterPro: IPR000536; Hormone_rec_1lg.
DR InterPro: IPR001628; Znf_C4steroid.
DR Pfam: PF00104; hormone_rec; 1.
DR Pfam: PF00105; zf-C4; 1.
DR ProDom: PD000035; Znf_C4steroid; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 448 AA; 50471 MW; BC8598A1D350701B CRC64;

Query Match 72.2%; Score 1760; DB 13; Length 448;
Best Local Similarity 74.6%; Pred. No. 3.1e-144;
Matches 343; Conservative 41; Mismatches 58; Indels 18; Gaps 5;

```

```
QY 1 MYSEVEGGTTPNFVVDYFNQ-N-RACILPEKGLPAGPYSTPLRPLM--NGSNHSIE 57
DB 1 MFCMDVLAVSPAQM--DYITASPSCMLQEKALKA---CFSGLAQTWOHRHSAQSVE 55
QY 58 TQSSSEIIVSPSPPLPRIVKPCFVCDKSSGYHYGSACGCKGFFRRSIOKNMY 117
DB 56 TQSTSEELVSPSPPLPRIVKPCFVCDKSSGYHYGSACGCKGFFRRSIOKNMY 115
QY 118 TCHRDKNCIINKYTRNPOCYCRLOKCFEVMGSKESVENDRNKKKKEVPKPECSESYTLTP 177
DB 116 TCHRDKNCIINKYTRNPOCYCRLOKCFEVMGSKESVENDRNKKKKEVPKPECSESYTLTP 175
QY 178 EVELTEKVKRAHQETFPALCOGLKYTTNNSSEQRVSLDIDLMDKFSLSKCIKTVEF 237
DB 176 ELDTLTKIRKAHQETFPALCOGLKYTTNNSADHRVLDLGLMDKFSLSKCIKTVEF 235
QY 238 AKOLPGFTTLTIADQITLLKAACLDILILRICRYTPEQDTMTFSDGLTLNRTOMHNAGF 297
DB 236 AKRLPGFTSLTIADQITLLKAACLDILILRICRYTPEQDTMTFSDGLTLNRTOMHNAGF 295
QY 298 GPLTDLVFAFANOLLPLEMDDAETGLLSAICLCGDRQDLEQDPRVDMLOEPLLEALKVY 357
DB 296 GPLTDLVFAFANOLLPLEMDDETGLLSAICLCGDRQDLEQDPRVDMLOEPLLEALKVY 355
QY 358 VKRRSRPHMFKMLKMTDLRSISAKGAERVITLMEIPGSMPLIQEMLENSEGLDT 417
DB 356 IRRRPNKPMFKMLKMTDLRSISAKGAERVITLMEIPGSMPLIQEMLENSEGLDT 415
QY 418 LSCOPCGGGRDGGGLPPPGSCSPSLSPSSNRSPATHSP 457
DB 416 LPTSTNGTAEH-----SPSISPSVDNSSVSQSP 445
RESULT 11
QY1YX2 PRELIMINARY; PRT; 458 AA.
AC QY1YX2;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Retinoic acid receptor, gamma.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; BC013709; AAH13709.1; -.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 458 AA; 50917 MW; B410D52805BE7CDD CRC64;
Query Match 67.7%; Score 1651; DB 11; Length 458;
Best Local Similarity 72.2%; Pred. No. 9.1e-135;
Matches 317; Conservative 40; Mismatches 50; Indels 32; Gaps 4;
QY 30 PEKGLPAGP-PYSTP-----LRTPLWNG-----SNHSIETQSSSEIIVP 68
DB 16 PGSGYFAGGFFAFPAFGALRGSPPFEMLSFSGRLGQDPLPKEMASLSVETQSTSEEMVP 75
QY 69 SPPSPPLPRIVKPCFVCDKSSGYHYGSACGCKGFFRRSIOKNMYTCHRDKNCIIN 128
DB 76 SPPSPPLPRIVKPCFVCDKSSGYHYGSACGCKGFFRRSIOKNMYTCHRDKNCIIN 135
QY 129 KVTNRNPOCYCRLOKCFEVMGSKESVENDRNKKKKEVPKPECSESYTLTPVEGELIEKVRK 188
DB 136 KVTNRNPOCYCRLOKCFEVMGSKESVENDRNKKKKEVPKPECSESYTLTPVEGELIEKVRK 195
```

```
DB 76 SSPSPPPPRIVKPCFVCDKSSGYHYGSACGCKGFFRRSIOKNMYTCHRDKNCIIN 135
QY 129 KVTNRNPOCYCRLOKCFEVMGSKESVENDRNKKKKEVPKPECSESYTLTPVEGELIEKVRK 188
DB 136 KVTNRNPOCYCRLOKCFEVMGSKESVENDRNKKKKEVPKPECSESYTLTPVEGELIEKVRK 195
QY 189 AHQETFPALCOGLKYTTNNSSEQRVSLDIDLMDKFSLSKCIKTVEFAKOLPGFTTLT 248
DB 196 AHQETFPALCOGLKYTTNNSADHRVLDLGLMDKFSLSKCIKTVEFAKOLPGFTTGLS 255
QY 249 IADQITLLKAACLDILILRICRYTPEQDTMTFSDGLTLNRTOMHNAGFPLTDLVFAFA 308
DB 256 IADQITLLKAACLDILILRICRYTPEQDTMTFSDGLTLNRTOMHNAGFPLTDLVFAFA 315
QY 309 NOLLPLEMDDAETGLLSAICLCGDRQDLEQDPRVDMLOEPLLEALKVYVRRRSPRPHM 368
DB 316 GOLLPLEMDDETGLLSAICLCGDRMDLEPEKVDKLOEPLLEALRYARRRRSQPYM 375
QY 369 FPKMLKITDLRSISAKGAERVITLMEIPGSMPLIQEMLENSEGLDTLSQPG----- 423
DB 376 FPKMLKITDLRGISTKGAERAITLMEIPGMPPLIREMLENPEMFEDDSKPGPHKA 435
QY 424 -----GGRDGGGLPPPP 436
DB 436 SSEDEAPGGGKGKQSGPOP 454
RESULT 12
QY1VX5 PRELIMINARY; PRT; 458 AA.
AC QY1VX5;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Similar to retinoic acid receptor, gamma.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-BREAST TUMOR;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL; BC012923; AAH12923.1; -.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW DNA-binding; Nuclear protein; Receptor; Transcription regulation;
KW Zinc-finger.
SQ SEQUENCE 458 AA; 50891 MW; 1D13D2F1482D7194 CRC64;
Query Match 67.7%; Score 1651; DB 11; Length 458;
Best Local Similarity 72.2%; Pred. No. 9.1e-135;
Matches 317; Conservative 40; Mismatches 50; Indels 32; Gaps 4;
QY 30 PEKGLPAGP-PYSTP-----LRTPLWNG-----SNHSIETQSSSEIIVP 68
DB 16 PGSGYFAGGFFAFPAFGALRGSPPFEMLSFSGRLGQDPLPKEMASLSVETQSTSEEMVP 75
QY 69 SPPSPPLPRIVKPCFVCDKSSGYHYGSACGCKGFFRRSIOKNMYTCHRDKNCIIN 128
DB 76 SPPSPPLPRIVKPCFVCDKSSGYHYGSACGCKGFFRRSIOKNMYTCHRDKNCIIN 135
QY 129 KVTNRNPOCYCRLOKCFEVMGSKESVENDRNKKKKEVPKPECSESYTLTPVEGELIEKVRK 188
DB 136 KVTNRNPOCYCRLOKCFEVMGSKESVENDRNKKKKEVPKPECSESYTLTPVEGELIEKVRK 195
```



```
QY 189 AHOETFPALCOLGKYTTNNSRQVSLDLDLNDKFSLSKTKCIKTVEFAKQLPGFTTIL 248
DB 196 AHOETFPSCQGLGKYTTNNSRQVSLDLDLNDKFSLSKTKCIKTVEFAKRLPGFTGLS 255
QY 249 IADQITLLKAACLDLILRICRYTPEQDTMTFSGLTLNRTOMHNGAGPLTDLVFAPA 308
DB 256 IADQITLLKAACLDLILRICRYTPEQDTMTFSGLTLNRTOMHNGAGPLTDLVFAPA 315
QY 309 NOLLPLEMDAETGLLSAICLCGRDQDLEQDPRVDMLEQPLEALKVYVRRRPSRPHM 368
DB 316 GOLLPLEMDTETGLLSAICLCGRDMDLEEKVKDQLQEPLEALRYARRRRSOPYM 375
QY 369 PPKMLKTLDRSISAKGAERVITLKMETPGSMPLIOEMLENSGLDLSQPG----- 423
DB 376 PPKMLKTLDRSISAKGAERVITLKMETPGSMPLIOEMLENSGLDLSQPG----- 423
QY 424 -----GGGRDGGGLPPPP 436
DB 436 SDEAPGCGQKRGQSPQ 454

RESULT 13
QY 08VHB6 PRELIMINARY; PRT; 443 AA.
AC 08VHB6;
DT 01-MAR-2002 (TrEMBLrel. 20, Created)
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Retinoic acid receptor gamma-2.
OS Mesocricetus auratus (Golden hamster).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
OC Mesocricetus.
OX NCBI_TaxID=10036;
RN [1]
RP SEQUENCE FROM N.A.
RA Sharma R.P., McGraw R.A., Dugyala R.R.;
RT "Cloning and characterization of hamster fetal retinoic acid receptor
RT isoforms."
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AY046945; AL023621;
DR InterPro: IPR000536; Hormone_rec_llg.
DR InterPro: IPR001723; Stdhrmn_receptor.
DR InterPro: IPR001628; Znf_C4steroid.
DR Pfam: PF00104; hormone_rec; 1.
DR Pfam: PF00105; zf-C4; 1.
DR PRINTS: PR00398; STRDHORMONER.
DR PRINTS: PR00047; STROIDFINGER.
DR ProDom: PD000035; Znf_C4steroid; 1.
DR SMART: SM00430; HOLI; 1.
DR SMART: SM00399; Znf_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; UNKNOWN_1.
KW Receptor.
SQ SEQUENCE 443 AA; 49431 MW; 36D86F26CBC4AB49 CRC64;

Query Match 67.6%; Score 1646.5; DB 11; Length 443;
Best Local Similarity 79.3%; Pred. No. 2.1e-134;
Matches 314; Conservative 32; Mismatches 37; Indels 13; Gaps 3;

QY 52 SNHSIETQSSSEIIVPSPPLPRIYKPCFCODKSSGYHYGVSAECGCKGFFRRSI 111
DB 48 SLOSIVETQSTSSSEIIVPSPPLPRIYKPCFCVNCNKKSSGYHYGVSSCEGCKGFFRSI 107
QY 112 QKNWYVTCRDKNCIINKVTRNFCOYCRLOKCFEYGVMSKSVNRNDRNKKKEVPKPCS- 170
DB 108 QKNWYVTCRDKNCIINKVTRNFCOYCRLOKCFEYGVMSKSVNRNDRNKKKEV-KEBCSP 166
QY 171 ESYTLTPEVGLIEKVRKAHQETFPALCOLGKYTTNNSRQVSLDLDLNDKFSLSKTK 230
DB 167 DSYELSPQLEELTKYSKAHOETFPALCOLGKYTTNNSRQVSLDLDLNDKFSLSKTK 226
QY 231 IIKTVFAKOLPGFTTILADQITLLKAACLDLILRICRYTPEQDTMTFSGLTLNRT 290
DB 231 IIKTVFAKOLPGFTTILADQITLLKAACLDLILRICRYTPEQDTMTFSGLTLNRT 290
```

```
DB 227 IIKVIFAKRLPGFTTGLSIADQITLLKAACLDLILMLRICRYTPEQDTMTFSGLTLNRT 286
QY 291 QHNAGFGPLTDLVFAPAFANOLLPLEMDAETGLLSAICLCGRDQDLEQDPRVDMLEQPL 350
DB 287 QHNAGFGPLTDLVFAPAGOLLPLEMDTETGLLSAICLCGRDMDLEEKVKDQLQEP 346
QY 351 LEALKVYVRRRPSRPHMFKMLKTLDRSISAKGAERVITLKMETPGSMPLIOEMLE 410
DB 347 LEALRYARRRRSOPYMFKMLKTLDRSISAKGAERVITLKMETPGSMPLIOEMLE 406
QY 411 NSEGLDLSQPG-----GGGRDGGGLPPP 435
DB 407 NPEMFEDDSKSGPHPKASSEDETPGGQKVGKLP 442

RESULT 14
QY 0918T3 PRELIMINARY; PRT; 441 AA.
AC 0918T3;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
DE Retonoic acid receptor gamma 2.
GN NR1B3B.
OS Ambystoma mexicanum (Axolotl).
OC Plasmid arARgamma2.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Caudata; Salamandroidae; Ambystomatidae;
OC Ambystoma.
OX NCBI_TaxID=8296;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRYO;
RX MEDLINE=20461762; PubMed=11004482;
RA Wirntan L., Seguin C.;
RT "Cloning of cDNAs encoding retinoic acid receptors RARgamma1,
RT RARgamma2, and a new splicing variant, RARgamma3, from Ambystoma
RT mexicanum and characterization of their expression during early
RT development."
RL Blochm. Biophys. Acta 1492:81-93(2000).
CC -|- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC -|- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY.
DR EMBL: AF206315; AAF80975.1;
DR HSP; P22932; 3LBD.
DR InterPro: IPR000536; Hormone_rec_llg.
DR InterPro: IPR001723; Stdhrmn_receptor.
DR InterPro: IPR001628; Znf_C4steroid.
DR Pfam: PF00104; hormone_rec; 1.
DR Pfam: PF00105; zf-C4; 1.
DR PRINTS: PR00398; STRDHORMONER.
DR PRINTS: PR00047; STROIDFINGER.
DR ProDom: PD000035; Znf_C4steroid; 1.
DR SMART: SM00430; HOLI; 1.
DR SMART: SM00399; Znf_C4; 1.
DR PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
KW DNA-binding; Nuclear protein; Plasmid; Receptor;
KW Transcription regulation; Zinc-finger.
SQ SEQUENCE 441 AA; 49865 MW; 906E25E145CEESE CRC64;

Query Match 67.3%; Score 1641; DB 13; Length 441;
Best Local Similarity 74.8%; Pred. No. 6.4e-134;
Matches 313; Conservative 37; Mismatches 56; Indels 14; Gaps 6;

QY 1 MYESVEGGTTPNPLVVDVFNQNRACILPEKGL-PAPG---PYSTPLRTPLWNGSNHSI 56
DB 1 MYDCMEAFMLAPHP--LYDVTNPG-ACMLRKLSPCFGLDPFGWFPASL-----QSV 52
QY 57 ETQSSSEIIVPSPPLPRIYKPCFCVODKSSGYHYGVSAECGCKGFFRRSTOKNV 116
DB 53 ETQSTSEEMVPSPPPPRVYKPCFCVNCNKKSSGYHYGVSSCEGCKGFFRRSTOKNV 112
QY 117 YTCRDKNCIINKVTRNFCOYCRLOKCFEYGVMSKSVNRNDRNKKKEVPKPCSSEYTLT 176
DB 117 YTCRDKNCIINKVTRNFCOYCRLOKCFEYGVMSKSVNRNDRNKKKEVPKPCSSEYTLT 176
```

Search completed: March 29, 2003, 07:57:21  
Job time : 65 secs

```

Query Match      66.8%; Score 1628.5; DB 13; Length 446;
Best Local Similarity 74.0%; Pred. No. 7.9e-133;
Matches 318; Conservative 37; Mismatches 56; Indels 19; Gaps 7;

Qy    1 MYTESVEVGPPNPPELVYDFNQNRACLLPEKGL-PAPG---PKSTPLRTPLWGNSHSI 56
      - - - - -
```

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: March 29, 2003, 05:12:23 ; Search time 23 Seconds  
(without alignments)

824.117 Million cell updates/sec

Title: US-09-691-220-2

Perfect score: 2437

Sequence: 1 MYSEVEGPTPNPLVDF.....SCSPSLSPSSNRSSPATHSP 457

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2127	87.3	462	1 RRA_HUMAN	P10276 homo sapien
2	2089	85.7	462	1 RRA_MOUSE	P11416 mus musculu
3	2020.5	82.9	460	1 RRA_CHICK	Q90966 gallus gall
4	1972	80.9	458	1 RRA_NOTVI	P18514 notophthalm
5	1895	77.8	458	1 RRA_XENLA	P51126 xenopus lae
6	1868.5	76.7	447	1 RRA_FUGRU	Q9W323 fugu rubrip
7	1761	72.3	455	1 RRB_HUMAN	P10826 homo sapien
8	1753.5	72.0	482	1 RRB_MOUSE	P22605 mus musculu
9	1736	71.2	455	1 RRB_COTJA	Q9W6b3 coturnix co
10	1733	71.1	455	1 RRB_CHICK	P22448 gallus gall
11	1663.5	68.3	499	1 RRG_BRARE	Q91392 brachydanio
12	1662	68.2	454	1 RRG1_HUMAN	P13631 homo sapien
13	1655.5	67.9	443	1 RRG2_HUMAN	P22932 homo sapien
14	1642	67.4	458	1 RRG1_MOUSE	P18911 mus musculu
15	1633.5	67.0	447	1 RRG2_MOUSE	P20787 mus musculu
16	1627.5	66.8	442	1 RRG2_XENLA	P28699 xenopus lae
17	1621	66.5	505	1 RRG_NOTVI	P18516 notophthalm
18	1620.5	66.5	476	1 RRG1_XENLA	P51127 xenopus lae
19	724	29.7	158	1 RRB_NOTVI	P18515 notophthalm
20	676.5	27.8	427	1 THA1_BRARE	Q98867 brachydanio
21	659.5	27.1	418	1 THA1_XENLA	P15204 xenopus lae
22	658	27.0	418	1 THA1_XENLA	P18115 xenopus lae
23	657.5	27.0	416	1 THA1_SALSA	Q9W785 salmo salar
24	653	26.8	410	1 THA1_SHEEP	Q28570 ovis aries
25	647	26.5	416	1 THA1_HIPHI	Q9W6n4 hippoglossu
26	647	26.5	418	1 THA1_RANCA	Q02777 rana catesb
27	646.5	26.5	391	1 THA1_PAROL	Q91242 paralichthy
28	646	26.5	416	1 THA1_PAROL	Q91241 paralichthy
29	645.5	26.5	408	1 THA1_CHICK	P04625 gallus gall
30	639	26.2	373	1 THB_RANCA	Q02965 rana catesb
31	634.5	26.0	401	1 THA_CAIMO	Q90382 cairina mos
32	630.5	25.9	395	1 THB_PAROL	Q91279 paralichthy
33	630	25.9	373	1 THBA_XENLA	P18117 xenopus lae

34	629.5	25.8	402	1 THA_APTPA	O42295 aptenodytes
35	629	25.8	386	1 THB_BRARE	Q9pve4 brachydanio
36	625.5	25.7	475	1 THB2_MOUSE	P37244 mus musculu
37	625	25.6	414	1 THBB_XENLA	P18119 xenopus lae
38	622.5	25.5	461	1 THB1_HUMAN	P10828 homo sapien
39	619	25.4	514	1 THB2_RAT	P37826 rattus norv
40	618	25.4	369	1 THB2_CHICK	P18112 gallus gall
41	617	25.3	476	1 THB2_HUMAN	P37243 homo sapien
42	615	25.2	461	1 THB1_MOUSE	P37242 mus musculu
43	615	25.2	461	1 THB1_RAT	P18113 rattus norv
44	607.5	24.9	402	1 THA_PYGAD	O42450 pygoscelis
45	606.5	24.9	385	1 ERBA_AVIER	P03373 avian eryth

ALIGNMENTS

RESULT 1

RRA_HUMAN	STANDARD;	PRT:	462 RA.
AC	P10276; Q13440; Q13441; P78456; Q9NQS0;		
DT	01-MAR-1989 (Rel. 10, Created)		
DT	01-OCT-1989 (Rel. 12, Last sequence update)		
DT	15-JUN-2002 (Rel. 41, Last annotation update)		
DE	Retinoic acid receptor alpha (RAR-alpha).		
GN	RARA OR NR1B1.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	[1]		
RP	SEQUENCE FROM N.A. (ISOFORM ALPHA-1).		
RX	MEDLINE=88055922; PubMed=2825036;		
RA	Giguere V., Ong E.S., Segui P., Evans R.M.;		
RT	"Identification of a receptor for the morphogen retinoic acid.";		
RL	Nature 330:624-629(1987).		
RN	[2]		
RP	SEQUENCE FROM N.A. (ISOFORM ALPHA-1).		
RX	MEDLINE=94134721; PubMed=8302850;		
RA	Chen Z., Guidex F., Rousselot P., Agadir A., Chen S.-J., Wang Z.-Y.,		
RT	Degos L., Zelent A., Waxman S., Chomienne C.;		
RT	"PLZF-RAR alpha fusion proteins generated from the variant		
RT	t(11;17)(q23;q21) translocation in acute promyelocytic leukemia		
RT	inhibit ligand-dependent transactivation of wild-type retinoic acid		
RT	receptors.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 91:1178-1182(1994).		
RN	[3]		
RP	SEQUENCE FROM N.A. (ISOFORM ALPHA-1).		
RX	MEDLINE=99269925; PubMed=10337631;		
RA	Hjalt T.A.H., Murray J.C.;		
RT	"Genomic structure of the human retinoic acid receptor-alpha gene.";		
RL	Mamm. Genome 10:528-529(1999).		
RN	[4]		
RP	SEQUENCE FROM N.A. (ISOFORM ALPHA-1).		
RX	TISSUE=Brain;		
RA	Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.		
RN	[5]		
RP	SEQUENCE OF 1-80 FROM N.A. (ISOFORM ALPHA-1).		
RX	MEDLINE=91088249; PubMed=2175878;		
RA	Brand N.J., Petkovich M., Chambon P.;		
RT	"Characterization of a functional promoter for the human retinoic acid		
RT	receptor-alpha (hRAR-alpha).";		
RL	Nucleic Acids Res. 18:6799-6806(1990).		
RN	[6]		
RP	SEQUENCE OF 31-462 FROM N.A.		
RX	MEDLINE=88055872; PubMed=2825025;		
RA	Petkovich M., Brand N.J., Krust A., Chambon P.;		
RT	"A human retinoic acid receptor which belongs to the family of nuclear		
RT	receptors.";		
RL	Nature 330:444-450(1987).		
RN	[7]		
RP	REVISIONS.		

RA Chabon P.;  
RN Submitted (DEC-1988) to the EMBL/GenBank/DBJ databases.  
RL [8]  
RP SEQUENCE OF 61-462 FROM N.A.  
RC TISSUE-Bone marrow;  
RX MEDLINE-96151966; PubMed-8562957;  
RA Redner R.L., Rush E.A., Faas S., Rudert W.A., Corey S.J.;  
RT "The t(5;17) variant of acute promyelocytic leukemia expresses a  
RL nucleophosmin-retinoic acid receptor fusion.";  
RN Blood 87:882-886(1996).  
RP SEQUENCE OF 1-54 FROM N.A. (ISOFORM ALPHA-2).  
RA Chen A., Petrie K., Waxman S., Zelent A.;  
RT "Homo sapiens retinoic acid receptor alpha (RAR-alpha) gene, promoter  
and 5' region of RAR-alpha 2 isoform.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE  
CC HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS  
CC A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS  
CC CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.  
CC -!- SUBCELLULAR LOCATION: Nuclear.  
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; ALPHA-1 (SHOWN HERE) AND ALPHA-  
CC 2; ARE PRODUCED BY ALTERNATIVE SPLICING. THEY DIFFER ONLY IN THEIR  
CC N-TERMINAL REGIONS.  
CC -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
CC -!- DISEASE: FORMS OF ACUTE PROMYELOCYTIC LEUKEMIA (APL) ARE  
CC CHARACTERIZED BY CHROMOSOMAL TRANSLOCATIONS T(11;17)(Q32;Q21)  
CC WHICH INVOLVES RARA AND PLZF, T(15;17)(Q21;Q21) WHICH INVOLVES  
CC RARA AND PML AND T(5;17)(Q32;Q11) THAT INVOLVES RARA AND NPM.  
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.  
CC NR1 SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X06614; CAA29829.1; -  
DR EMBL; X06538; CAA29787.1; -  
DR EMBL; AF088895; AAD05222.1; -  
DR EMBL; AF088889; AAD05222.1; JOINED.  
DR EMBL; AF088890; AAD05222.1; JOINED.  
DR EMBL; AF088891; AAD05222.1; JOINED.  
DR EMBL; AF088892; AAD05222.1; JOINED.  
DR EMBL; AF088893; AAD05222.1; JOINED.  
DR EMBL; AF088894; AAD05222.1; JOINED.  
DR EMBL; BC008727; AAO8727.1; -  
DR EMBL; X56058; CAA39533.1; -  
DR EMBL; X56685; CAA39533.1; JOINED.  
DR EMBL; X58685; CAA41532.1; -  
DR EMBL; U41742; AAB00112.1; ALT-INIT.  
DR EMBL; U41743; AAB00113.1; ALT-INIT.  
DR EMBL; AF283809; AAF87249.1; -  
DR PIR; A29491; A29491.  
DR PIR; S07272; S07272.  
DR HSSP; P10826; 1HRA.  
DR TRANSFAC; T00719; -  
DR Genew; HGNC:9864; RARA.  
DR MIM; 180240; -  
DR InterPro; IPR000536; Hormone\_rec\_lig.  
DR InterPro; IPR001723; Stdrhm\_receptor.  
DR InterPro; IPR001628; znf\_C4steroid.  
DR Pfam; PF00104; hormone\_rec; 1.  
DR Pfam; PF00105; zf-C4; 1.  
DR PRINTS; PR00398; STRDHORMONER.  
DR PRINTS; PR00047; STROIDFINGER.  
DR ProDom; PD000035; znf\_C4steroid; 1.  
DR SMART; SM00430; HOL1; 1.  
DR SMART; SM00399; znf\_C4; 1.

DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
KW Zinc-finger; Multigene family; Alternative splicing; Proto-oncogene;  
KW Chromosomal translocation.  
FT DOMAIN 1 87 MODULATING.  
FT DNA\_BIND 88 153 NUCLEAR RECEPTOR-TYPE.  
FT ZN\_FING 88 108 C4-TYPE.  
FT ZN\_FING 124 148 C4-TYPE.  
FT DOMAIN 154 199 HINGE.  
FT DOMAIN 200 419 LIGAND-BINDING.  
FT SITE 60 61 BREAKPOINT FOR TRANSLOCATION TO FORM  
FT PLZF-RAR-ALPHA AND RAR-ALPHA-PLZF  
FT ONCOGENES.  
FT VARSPLIC 1 60 MASNSSCPTGGGHLNCPYPPVAFPPMGLGSLPPGAL  
FT TTHQLPLVSGYSTSPAT -> MIESVEVGPTNPPLV  
FT DFYNRACLLPEKGLPAPGPYSTPLRTPLWNGSNHS (IN  
FT ISOFORM ALPHA-2).  
FT CONFLICT 241 241 E -> D (IN REF. 3).  
SQ SEQUENCE 462 AA; 50771 MW; E8D1CF9A1E57CB99 CRC64;  
Query Match 87.3%; Score 2127; DB 1; Length 462;  
Best Local Similarity 91.6%; Pred. No. 2.5e-149;  
Matches 413; Conservative 5; Mismatches 13; Indels 20; Gaps 4;  
QY 7 VGGTPNPFLVDFYNQNRACLLPEKGLPAPGPYSTPLRTPLWNGSNHSIETQSSSSEI 66  
DB 32 LGGSP-PGALTTLQHQ-----LPVSG-YSTP-----SPATIEIQTSSSSEI 71  
QY 67 VPSPSPPLPRIYKPCFVCODKSGYHYGVSAECGCKGFFRRSIQKNVYTCRDKNCI 126  
DB 72 VPSPSPPLPRIYKPCFVCODKSGYHYGVSAECGCKGFFRRSIQKNVYTCRDKNCI 131  
QY 127 INKVRNCPQYCRLOKCFEVMGSKESVRNDRNKKKVKPECSYSLTPEVGELIEKV 186  
DB 132 INKVRNCPQYCRLOKCFEVMGSKESVRNDRNKKKVKPECSYSLTPEVGELIEKV 191  
QY 187 RKAHOETFPALCOLGKYTTNNSQRYSLDLDLWDKSELSKCIKTVEAKOLPGFTT 246  
DB 192 RKAHOETFPALCOLGKYTTNNSQRYSLDLDLWDKSELSKCIKTVEAKOLPGFTT 251  
QY 247 LTIADQITLLKACALDILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFPLTDLVFA 306  
DB 252 LTIADQITLLKACALDILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFPLTDLVFA 311  
QY 307 FANQLPLEMDDAETGLLSAICLCGDRQDLQEPDRVDMLOEPLLEALKVYVRRRSPRP 366  
DB 312 FANQLPLEMDDAETGLLSAICLCGDRQDLQEPDRVDMLOEPLLEALKVYVRRRSPRP 371  
QY 367 HMPFKMLKITDLRSISAKGAERVITLKMETPGSMPPLIQEMLENSEGLDLSGPGGGG 426  
DB 372 HMPFKMLKITDLRSISAKGAERVITLKMETPGSMPPLIQEMLENSEGLDLSGPGGGG 431  
QY 427 RDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457  
DB 432 RDGGGLAPPGSCSPSLSPSSNRSSPATHSP 462  
RESULT 2  
RRA\_MOUSE  
ID RRA\_MOUSE STANDARD; PRT; 462 AA.  
AC P11416; P22603;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Retinoic acid receptor alpha (RAR-alpha).  
GN RARA OR NR1B1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Euthera; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_Taxid=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM ALPHA-1).  
RX MEDLINE-89295563; PubMed-2544807;

RA Zelent A., Krust A., Petkovitch M., Kastner P., Chambon P.:  
 RT "Cloning of murine alpha and beta retinoic acid receptors and a novel  
 RL receptor gamma predominantly expressed in skin.";  
 RN Nature 339:714-717(1989).  
 RP [2]  
 RX SEQUENCE FROM N.A. (ISOFORM ALPHA-1).  
 RA MEDLINE-93195858; PubMed-8383767;  
 RA Heilmann R., Rentrop M., Lang E., Maelicke A.;  
 RT "Cloning of several genes coding for retinoic acid nuclear receptors  
 RL in the mouse embryonal carcinoma cell line PCC7-MZ1.";  
 RN J. Recept. Res. 13:693-709(1993).  
 RP [3]  
 RX SEQUENCE FROM N.A. (ISOFORMS ALPHA-1 AND ALPHA-2).  
 RA MEDLINE-91114713; PubMed-1846598;  
 RA Leroy P., Krust A., Zelent A., Mendelsohn C., Garnier J.-M.,  
 RA Kastner P., Dierich A., Chambon P.;  
 RT "Multiple isoforms of the mouse retinoic acid receptor alpha are  
 RL generated by alternative splicing and differential induction by  
 RL retinoic acid.";  
 RN EMBO J. 10:59-69(1991).  
 RP [4]  
 RX SEQUENCE FROM N.A. (VARIANT IN EMBRYONAL CARCINOMA CELL LINE RAC65).  
 RA MEDLINE-92324516; PubMed-1320576;  
 RA Kruijt F.A.E., van der Veer L., Mader S., van den Brink C.E.,  
 RA Feijen A., Jonk L.J., Kruijer W., van der Saag P.T.;  
 RT "Retinoic acid resistance of the variant embryonal carcinoma cell  
 RL line RAC65 is caused by expression of a truncated RAR alpha.";  
 RN Differentiation 49:27-37(1992).  
 RP [5]  
 RX SEQUENCE FROM N.A. (VARIANT IN EMBRYONAL CARCINOMA CELL LINE RAC65).  
 RA MEDLINE-91061752; PubMed-2174108;  
 RA Pratt M.A.C., Kralova J., McBurney M.W.;  
 RT "A dominant negative mutation of the alpha retinoic acid receptor  
 RL gene in a retinoic acid-nonresponsive embryonal carcinoma cell.";  
 RN Mol. Cell. Biol. 10:6445-6453(1990).  
 CC -1- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE  
 CC HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS  
 CC A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS  
 CC CELL FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.  
 CC -1- SUBCELLULAR LOCATION: Nuclear.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: ALPHA-1 (SHOWN HERE) AND ALPHA-  
 CC 2; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
 CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
 CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.  
 CC NR1 SUBFAMILY.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X56572; CAA39919.1; -  
 CC EMBL; X56565; CAA39917.1; -  
 CC EMBL; S56656; AAB25783.1; -  
 CC EMBL; X57528; CAA40749.1; -  
 CC EMBL; M60909; AAA40031.1; -  
 CC F01; S05050; S05050.  
 CC PIR; S13599; S13599.  
 CC HSP; P10826; 1HRA.  
 CC TRANSFAC; T00717; -  
 CC TRANSFAC; T01327; -  
 CC TRANSFAC; T01335; -  
 CC TRANSFAC; T01336; -  
 CC MGD; MGI:97856; Rara.  
 CC InterPro; IPR000536; Hormone\_rec\_lig.  
 CC InterPro; IPR001723; Steroid\_receptor.  
 CC InterPro; IPR001628; Znf\_C4steroid.  
 CC Pfam; PF00104; hormone\_rec; 1.  
 CC Pfam; PF00105; zf-C4; 1.

DR PRINTS; PR00398; STERHORMONER.  
 DR PRINTS; PR00047; STROIDFINGER.  
 DR PRODOM; PD000035; Znf\_C4steroid; 1.  
 DR SMART; SM00430; HOL1; 1.  
 DR SMART; SM00399; Znf\_C4; 1.  
 DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
 KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
 RL Zinc-finger; Multigene family; Alternative splicing.  
 FT DOMAIN 1 87 MODULATING.  
 FT DNA\_BIND 88 153 NUCLEAR RECEPTOR-TYPE.  
 FT ZN\_FING 88 108 C4-TYPE.  
 FT ZN\_FING 124 148 C4-TYPE.  
 FT DOMAIN 154 199 HINGE.  
 FT DOMAIN 200 419 LIGAND-BINDING.  
 FT VARSPLIC 1 60  
 FT VARIANT 391 391 G -> A (IN EMBRYONAL CARCINOMA CELL LINE  
 FT RAC65).  
 FT VARIANT 392 462 MISSING (IN EMBRYONAL CARCINOMA CELL LINE  
 FT RAC65).  
 FT CONFLICT 163 163 N -> K (IN REF. 5).  
 FT CONFLICT 179 179 T -> S (IN REF. 5).  
 FT CONFLICT 284 284 M -> L (IN REF. 5).  
 SQ SEQUENCE 462 AA; 50735 MW; 726F77996338A5AD CRC64;  
 Query Match 85.7%; Score 2089; DB 1; Length 462;  
 Best Local Similarity 90.0%; Pred. No. 1.6e-146;  
 Matches 406; Conservative 8; Mismatches 17; Indels 20; Gaps 4;  
 QY 7 VGGTPNPFLVVDYNNRACLLPEKGLPAGPYSTPLRTPLWNGSNHSIETQSSSEI 66  
 DB 32 LGGLSLPGALTSLQHQ-----LPVSG-YSTP-----SPATIEQSSSEI 71  
 QY 67 VPSPSPPLPRIYKPCFVCDKSSGVHYGVSAEGCKGFFRRSIQKNVYTCHRDKNCI 126  
 DB 72 VPSPSPPLPRIYKPCFVCDKSSGVHYGVSAEGCKGFFRRSIQKNVYTCHRDKNCI 131  
 QY 127 INKVTNRPCQYCRLOKCFEVMGSKESVNRDNKKKKKPKCESESYTLTPVGELEKV 186  
 DB 132 INKVTNRPCQYCRLOKCFEVMGSKESVNRDNKKKKKPKCESESYTLTPVGELEKV 191  
 QY 187 RKAHQETFPALCQGLKYTTNNSSQVSLDIDLMDKFSLSKCIKTVFAKQLPGFTT 246  
 DB 192 RKAHQETFPALCQGLKYTTNNSSQVSLDIDLMDKFSLSKCIKTVFAKQLPGFTT 251  
 QY 247 LTIADQITLLKAACLDILILICITRYTPEQDTMTFSDGLTLNRTOMHNAFGPLTDLVFA 306  
 DB 252 LTIADQITLLKAACLDILILICITRYTPEQDTMTFSDGLTLNRTOMHNAFGPLTDLVFA 311  
 QY 307 FANOLLPLEMDDAETGLLSAICLCIGDRQDLQEPDRVMDLQEPLEALKVYVRRRSPR 366  
 DB 312 FANOLLPLEMDDAETGLLSAICLCIGDRQDLQEPDRVMDLQEPLEALKVYVRRRSPR 371  
 QY 367 HMFPEKMLKMTDLRSISAKGAERVITLKMETPGSMPLIQEMLENSEGLDTLSGPGGGG 426  
 DB 372 HMFPEKMLKMTDLRSISAKGAERVITLKMETPGSMPLIQEMLENSEGLDTLSGPGGGG 431  
 QY 427 RDGGGLPPPGSCSPSLSPSSNRSPATHSP 457  
 DB 432 RDGGGLAPPPGSCSPSLSPSSNRSPATHSP 462  
 RESULT 3  
 ID RRA\_CHICK STANDARD; PRT; 460 AA.  
 AC Q90966; Q90967;  
 DT 15-JUN-2002 (Rel. 41, Created)  
 DT 15-JUN-2002 (Rel. 41, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Retinoic acid receptor alpha (RAR-alpha).  
 GN RARA OR NR1B1.



```

DR EMBL; X17585; CAA35602.1; -
DR EMBL; Z14254; CAA78621.1; -
DR PIR; S06123; S06123.
DR HSP; P10826; 1HRA.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhrmn_receptor.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 2.
DR Pfam; PF00105; zf-C4; 2.
DR PRINTS; PR00398; STRDHOMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Multigene family; Alternative splicing.
FT DOMAIN 1 86
FT DNA_BIND 87 152 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 87 107 C4-TYPE.
FT ZN_FING 123 147 C4-TYPE.
FT ZN_FING 153 199 HINGE.
FT DOMAIN 200 419 LIGAND-BINDING.
FT VARSPLIC 1 59 MASNGSCPPSSGGHNGYVPVHYAFFPHMLGGLSPGSLA
FT GIPHPVSAISTPAT -> MDSVEVSPSPYIMIDFY
FT SQACLMADKGLGHPVPGSPPIRPHWSSSHS (IN
FT ISOFORM ALPHA-2).
SQ SEQUENCE 458 AA; 50637 MW; B4370822FBAADB54 CRC64;

Query Match 80.98; Score 1972; DB 1; Length 458;
Best Local Similarity 88.3%; Pred. No. 6.4e-138;
Matches 379; Conservative 16; Mismatches 18; Indels 16; Gaps 4;

QY 33 GLPAPGP---YSTPLRTPRWNGSHSIETSSSEIIVSPSPPLPRIYKPCFVCODK 89
DB 42 GIPHPVSAISTP-----SPATIEQTSSSEIIVSPSPPLPRIYKPCFVCODK 93

QY 90 SSGHYGVSAEGCGKGFRRSIQKNMYTCHRDKNCIINKVTRNRCQYCRQKCFEVMGS 149
DB 94 SSGHYGVSAEGCGKGFRRSIQKNMYTCHRDKTCIINKVTRNRCQYCRQKCFEVMGS 153

QY 150 KESVNRDNRNKKK-EVPKPECSSEYTLTPVEGLIEKVRKAHQETFPALCOLGKYTTNNS 208
DB 154 KESVNRDNRNKKKQEPKQECTESIYITPEVEDLVEKVRKAHQETFPALCOLGKYTTNNS 213

QY 209 SEQRVSLDIDLWDFSELSTKCIITKVEFAKOLPGFTTLTIADQITLLKAAACLDLILRI 268
DB 214 SEERVSLDIDLWDFSELSTKCIITKVEFAKOLPGFTTLTIADQITLLKAAACLDLILRI 273

QY 269 CTRYTPQDTMTFSDGLTLNRTQMHNAGFGPLTDLVFAFANQLLPLEMDDAETGLLSAIC 328
DB 274 CTRYTPQDTMTFSDGLTLNRTQMHNAGFGPLTDLVFAFANQLLPLEMDDAETGLLSAIC 333

QY 329 LICGDRDLQEPQDVMDLQEPLEALKVYKRRPSPHPKMLKITDLSRSISAKGAE 388
DB 334 LICGDRDLQEPQDVMDLQEPLEALKVYKRRPSPHPKMLKITDLSRSISAKGAE 393

QY 389 RVTLKMEIPGSPPLIQEMLENSEGLDLSGQPGCGGRDGGGLPPPGSCSPSLSPSSN 448
DB 394 RVTLKMEIPGSPPLIQEMLENSEGLDLSGTGP-----PRASSLAPPGGSCSPSLSPSSN 449

QY 449 RSSPATHSP 457
DB 450 RSSPTSHSP 458

RESULT 5
RRA_XENLA STANDARD; PRT; 458 AA.
ID RRA_XENLA
AC P51126.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

Retinoic acid receptor alpha (RAR-alpha).
RARA OR NR1B1.
Xenopus laevis (African clawed frog).
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
Xenopodinae; Xenopus.
NCBI_TaxID=8355;
[1]
SEQUENCE FROM N.A.
MEDLINE=92196110; PubMed=1312717;
Blumberg B., Mangelsdorf D.J., Dyck J.A., Bittner D.A., Evans R.M.,
de Robertis E.M.;
"Multiple retinoid-responsive receptors in a single cell: families of
retinoid 'x' receptors and retinoic acid receptors in the Xenopus
egg.";
Proc. Natl. Acad. Sci. U.S.A. 89:2321-2325(1992).
CC - FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE
HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS
A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS
CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.
CC - SUBCELLULAR LOCATION: Nuclear.
CC - DEVELOPMENTAL STAGE: IT IS SYNTHESIZED DURING OOGENESIS AND
PERSISTS IN THE CLEAVING EMBRYO AT APPROXIMATELY CONSTANT LEVELS
UNTIL IT IS DEGRADED JUST BEFORE GASTRULATION.
CC - DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC - SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NR1 SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

EMBL; L11445; -; NOT_ANNOTATED_CDS.
DR HSP; P10826; 1HRA.
DR TRANSFAC; T01356.
DR InterPro; IPR000536; Hormone_rec_lig.
DR InterPro; IPR001723; Stdhrmn_receptor.
DR InterPro; IPR001628; Znf_C4steroid.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHOMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR ProDom; PD000035; Znf_C4steroid; 1.
DR SMART; SM00430; HOL1; 1.
DR SMART; SM00399; Znf_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger; Multigene family.
FT DOMAIN 1 87
FT DNA_BIND 88 153 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 88 108 C4-TYPE.
FT ZN_FING 124 148 C4-TYPE.
FT ZN_FING 154 199 HINGE.
FT DOMAIN 200 418 LIGAND-BINDING (BY SIMILARITY).
FT SEQUENCE 458 AA; 50573 MW; 4D80BB18678B1E17 CRC64;

Query Match 77.8%; Score 1895; DB 1; Length 458;
Best Local Similarity 80.5%; Pred. No. 3e-132;
Matches 372; Conservative 20; Mismatches 44; Indels 26; Gaps 5;

QY 10 PTPNFPVLVDYFNONRACLLPEK-----GLPAP-----GYPSTPLRTPRWNGSHS 55
DB 9 PPGPGHINGHFVHYAFFPHMLGCMSTGGLPGVQHPPLSGYSTP-----SPAT 60

QY 56 IETQSSSEIIVSPSPPLPRIYKPCFVCODKSSGGHYGVSAEGCKGFFRRSIQKNM 115
DB 61 IETQSTSEIIVSPSPPLPRIYKPCFVCODKSSGGHYGVSAEGCKGFFRRSIQKNM 120

```





CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OX Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM BETA-2).  
RC TISSUE=Placenta;  
RX MEDLINE=88232961; PubMed=2836738;  
RA Benbrook D., Lernherdt E., Pfahl M.;  
RT "A new retinoic acid receptor identified from a hepatocellular  
RL carcinoma.";  
RL Nature 333:669-672(1988).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM BETA-2).  
RX MEDLINE=88065931; PubMed=2825037;  
RA de The H., Marchio A., Tiollais P., Dejean A.;  
RT "A novel steroid thyroid hormone receptor-related gene  
RL inappropriately expressed in human hepatocellular carcinoma.";  
RL Nature 330:667-670(1987).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM BETA-4).  
RC TISSUE=Breast tumor;  
RX MEDLINE=99342073; PubMed=10411930;  
RA Sommer K.M., Chen L.I., Treuting P.M., Smith L.T., Swisshelm K.;  
RT "Elevated retinoic acid receptor beta(4) protein in human breast tumor  
RL cells with nuclear and cytoplasmic localization.";  
RL Proc. Natl. Acad. Sci. U.S.A. 96:8651-8656(1999).  
RN [4]  
RP SEQUENCE OF 1-72 FROM N.A. (ISOFORM BETA-1).  
RX MEDLINE=94101604; PubMed=8275470;  
RA Houle B., Pelletier M., Wu J., Goodyer C., Bradley W.E.;  
RT "Petal isoform of human retinoic acid receptor beta expressed in small  
RL cell lung cancer lines.";  
RL Cancer Res. 54:365-369(1994).  
RN [5]  
RP SEQUENCE OF 61-109 FROM N.A.  
RX MEDLINE=86257411; PubMed=3014347;  
RA Dejean A., Bougueleret L., Grzeschik K.-H., Tiollais P.;  
RT "Hepatitis B virus DNA integration in a sequence homologous to  
RL v-erbA and steroid receptor genes in a hepatocellular carcinoma.";  
RL Nature 322:70-72(1986).  
RN [6]  
RP SEQUENCE OF 61-109 FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=91014673; PubMed=2170809;  
RA Dejean A., de The H.;  
RT "Hepatitis B virus as an insertional mutagen in a human  
RL hepatocellular carcinoma.";  
RL Mol. Biol. Med. 7:213-222(1990).  
RN [7]  
RP IDENTITY OF HAP AND RAR-BETA.  
RX MEDLINE=88189352; PubMed=2833708;  
RA Brand N., Petkovitch M., Krust A., Chambon P., de The H., Marchio A.,  
RA Tiollais P., Dejean A.;  
RT "Identification of a second human retinoic acid receptor.";  
RL Nature 332:850-853(1988).  
RN [8]  
RP STRUCTURE BY NMR OF 82-160.  
RX MEDLINE=92338160; PubMed=1321662;  
RA Kathira M., Knegetel R.M.A., Boelens R., Elb D., Schilthuis J.G.,  
RA van der Saag P.T., Kaptein R.;  
RT "Homo- and heteronuclear NMR studies of the human retinoic acid  
RL receptor beta DNA-binding domain: sequential assignments and  
RL identification of secondary structure elements.";  
RL Biochemistry 31:6474-6480(1992).  
RN [9]  
RP STRUCTURE BY NMR OF 82-160.  
RX MEDLINE=93192652; PubMed=8383553;  
RA Knegetel R.M., Kathira M., Schilthuis J.G., Bonvin A.M., Boelens R.,  
RA Elb D., van der Saag P.T., Kaptein R.;  
RT "The solution structure of the human retinoic acid receptor-beta DNA-  
RL binding domain.";  
RL J. Biomol. NMR 3:1-17(1993).  
CC -1- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE

CC HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS  
CC A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS  
CC CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: Nuclear (isoforms beta-1 and beta-2) and  
CC cytoplasmic (isoform beta-4).  
CC -1- ALTERNATIVE PRODUCTS: 4 isoforms; beta-1 (shown here), beta-2,  
CC beta-3 and beta-4; are produced by alternative splicing.  
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.  
CC NR1 SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC -----  
DR EMBL; X07282; CAA30262.1; -  
DR EMBL; Y00291; CAA68398.1; -  
DR EMBL; AF157483; AAD45688.1; -  
DR EMBL; X77664; CAA54740.1; -  
DR EMBL; X04014; CAA27637.1; ALT\_SEQ.  
DR EMBL; M57445; AAA58728.1; -  
DR PIR; S02827; S02827.  
DR PIR; A29492; A29492.  
DR PIR; A25721; TVHUEA.  
DR PDB; 1HRA; 31-JAN-94.  
DR TRANSFAC; T00721; -  
DR TRANSFAC; T01326; -  
DR Gene; HGNC:9865; RARB.  
DR MIM; 180220; -  
DR InterPro; IPR000536; Hormone\_rec\_lig.  
DR InterPro; IPR001723; Steroid\_receptor.  
DR InterPro; IPR001628; Znf\_C4steroid.  
DR Pfam; PF00104; hormone\_rec; 2.  
DR Pfam; PF00105; zf-C4; 4.  
DR PRINTS; PR00398; STRDHORMONER.  
DR PROSITE; PR00047; STEROIDFINGER.  
DR PROSITE; PR000035; Znf\_C4steroid; 1.  
DR SMART; SM00430; HOLI; 2.  
DR SMART; SM00399; Znf\_C4; 2.  
DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
KW Zinc-finger; Multigene family; Alternative splicing; Proto-oncogene;  
KW 3D-structure.  
FT DOMAIN 1 87 MODULATING.  
FT DNA\_BIND 88 153 NUCLEAR RECEPTOR-TYPE.  
FT ZN\_FING 88 108 C4-TYPE.  
FT ZN\_FING 124 148 C4-TYPE.  
FT DOMAIN 154 199 HINGE.  
FT DOMAIN 200 419 LIGAND-BINDING.  
FT VARSPPLIC 1 60 MTTSGHACVPVAVNGHMYPATYPLLPPLPPVIGLSLPL  
HGLHGHPPSGGCTSPAT -> MFCMDVLSVSPGQILDF  
YTASPSMLOEKALKACFSLGTQTEWQHRHTAQS (IN  
ISOFORM BETA-2).  
FT VARSPPLIC 1 119 MISSING (IN ISOFORM BETA-4).  
FT CONFLICT 206 206 G -> A (IN REF. 2).  
FT CONFLICT 317 317 L -> Q (IN REF. 1).  
FT CONFLICT 414 414 L -> M (IN REF. 2).  
FT CONFLICT 454 454 V -> L (IN REF. 1).  
FT HELIX 107 118  
FT TURN 119 119  
FT TURN 128 129  
FT HELIX 141 151  
FT TURN 158 159  
SQ SEQUENCE 455 AA; 50489 MW; 8813263AD0495D5A CRC64;  
Query Match 72.3%; Score 1761; DB 1; Length 455;  
Best Local Similarity 79.1%; Pred. No. 2.2e-122;  
Matches 336; Conservative 30; Mismatches 41; Indels 18; Gaps 2;

```
QY 33 GLPAPGYPSTPLRTPLWNGSNHSIETQSSSEIVSPSPPLPRIYKPCFVCQDKSSG 92
DB 46 GHPPPGSGCTP-----SPATIEQTSSSEELVSPSPPLPPRVYKPCFVCQDKSSG 97
QY 93 YHGVGSACGCKGFFRRSIQKNVYTCRDKNCIINKVTRNCPQYCRLOKCFEFGVSKES 152
DB 98 YHGVGSACGCKGFFRRSIQKNVYTCRDKNCIINKVTRNCPQYCRLOKCFEFGVSKES 157
QY 153 VRNDRNKKKKVPKPCSSSYTLTPVEGBELIEKVRKHAQETFPALCQLGKYTTNNSSEOR 212
DB 158 VRNDRNKKKKETSKQCBCTSEYEMTAEUDDLTETKIRKHAQETFPALCQLGKYTTNNSADHR 217
QY 213 VSLDLDLWDFSELSTKCIKIIVFAKQLPGFTTLTIADQITLLKACLDLILRLICTRY 272
DB 218 VRLDGLMDKFSLATKCIKIIVFAKRLPGFTTLTIADQITLLKACLDLILRLICTRY 277
QY 273 TPEQDTMTFSDGLTLNRTOMHNGAGFGLDLVAFANQLLPLEMDDAETGLLSAICLIG 332
DB 278 TPEQDTMTFSDGLTLNRTOMHNGAGFGLDLVAFANQLLPLEMDDTETGLLSAICLIG 337
QY 333 DRQDLQPDVMDLQEPLEALKVYKRRSPRPHMFKMLMKITDLRSISAKGAERVIT 392
DB 338 DRQDLQPDVMDLQEPLEALKVYKRRSPRPHMFKMLMKITDLRSISAKGAERVIT 397
QY 393 LKMEIPGSMPLIQEMLNSEGDLTSLGQPGGGGGGLPPPPGSCPSLSPSSNRSSP 452
DB 398 LKMEIPGSMPLIQEMLNSEGHEPLTPSSSGNTAEH-----SPSISPVSSVENS 447
QY 453 ATHSP 457
DB 448 VSQSP 452

RESULT 8
RBL_MOUSE
ID RBL_MOUSE STANDARD; PRT: 482 AA.
AC P22605; P22604; P11417;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Retinoic acid receptor beta (RAR-beta).
GN RAR OR NR1B2.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS BETA-1; BETA-2 AND BETA-3).
RC STRAIN=C57BL/6;
RX MEDLINE=91114714; PubMed=1846599;
RA Zelen A., Mendelsohn C., Kastner P., Krust A., Garnier J.-M.,
RA Ruffenach F., Leroy P., Chambon P.;
RT "Differentially expressed isoforms of the mouse retinoic acid
RT receptor beta generated by usage of two promoters and alternative
RT splicing.";
RL EMBO J. 10:71-81(1991).
RN [2]
RP SEQUENCE FROM N.A. (ISOFORM BETA-2).
RX MEDLINE=89295563; PubMed=2544807;
RA Zelen A., Krust A., Petkovitch M., Kastner P., Chambon P.;
RT "Cloning of murine alpha and beta retinoic acid receptors and a novel
RT receptor gamma predominantly expressed in skin.";
RL Nature 339:714-717(1989).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM BETA-2).
RX MEDLINE=93195850; PubMed=8383767;
RA Heiermann R., Rentrop M., Lang E., Maelicke A.;
RT "Cloning of several genes coding for retinoic acid nuclear receptors
RT in the mouse embryonal carcinoma cell line PCC7-MZ1.";
RL J. Recept. Res. 13:693-709(1993).
RN [4]
RP SEQUENCE FROM N.A. (ISOFORM BETA-4).
```

```
RX MEDLINE=92212900; PubMed=1313565;
RA Nagpal S., Zelen A., Chambon P.;
RT "RAR-beta 4, a retinoic acid receptor isoform is generated from
RT RAR-beta 2 by alternative splicing and usage of a CUG initiator
RT codon.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:2718-2722(1992).
CC -1- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE
CC HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS
CC A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS
CC CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- ALTERNATIVE PRODUCTS: 4 ISOFORMS; BETA-1, BETA-2, BETA-3 (SHOWN
CC HERE) AND BETA-4; ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NRI SUBFAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X56574; CAA39921.1; -
CC EMBL: X56569; CAA39918.1; -
CC EMBL: X56573; CAA39920.1; -
CC EMBL: X56660; AAB25784.2; -
CC EMBL: S92180; -; NOT_ANNOTATED_CDS.
CC PIR: S13515; S13515
CC PIR: S13516; S13516
CC PIR: S13517; S13517
CC PIR: S05051; S05051
CC HSP: P10826; IHRA.
CC TRANSFAC: T00717; -
CC TRANSFAC: T01337; -
CC TRANSFAC: T01338; -
CC TRANSFAC: T01339; -
CC MGD: MGI:97857; Rarb.
CC InterPro: IPR000536; Hormone_rec_lig.
CC InterPro: IPR001723; Stdhrmn_receptor.
CC Pfam: PF00104; hormone_rec; 1.
CC Pfam: PF00105; zf-C4; 1.
CC PRINTS: PR00398; STRDHORMONER.
CC PRINTS: PR00047; STROIDFINGER.
CC ProDom: PD000035; Znf_C4steroid; 1.
CC SMART: SM00430; HOL1; 1.
CC SMART: SM00399; Znf_C4; 1.
CC PROSITE: PS00031; NUCLEAR_RECEPTOR; 1.
CC Receptor; Transcription regulation; DNA-binding; Nuclear protein;
CC Zinc-finger; Multigene family; Alternative splicing.
CC DOMAIN 1 114 MODULATING.
CC DNA_BIND 115 180 NUCLEAR RECEPTOR-TYPE.
CC ZN_FING 115 135 C4-TYPE.
CC ZN_FING 151 175 C4-TYPE.
CC DOMAIN 181 226 HINGE.
CC DOMAIN 227 446 LIGAND-BINDING.
CC VARSPLIC 1 87 MSTSHACPVAVRGHMTYPAAPYPLFPVIRGLSLPL
CC FT HGLHGHPPSPSCSPSPASVQACORTGGSOFAASTKWP
CC FT LKNA -> MFDCMDVLSVSGQILDFTYSPSCMLQEA
CC FT LKACLSGFTQAEWHRRHTAQS (IN ISOFORM
CC FT BETA-2).
CC FT MISSING (IN ISOFORM BETA-4).
CC FT LNA -> MEN (IN ISOFORM BETA-4).
CC FT S -> T (IN ISOFORM BETA-1).
CC FT MISSING (IN ISOFORM BETA-1).
CC FT VARSPLIC 61 87
CC FT VARSPLIC 482 AA; 53331 MW; 48E78E6C7D012515 CRC64;
CC FT SEQUENCE
CC
CC Query Match 72.0%; Score 1753.5; DB 1; Length 482;
CC Best Local Similarity 76.1%; Pred. No. 8.4e-122;
```

Matches 338; Conservative 31; Mismatches 46; Indels 29; Gaps 3;  
QY 33 GLPAPGYPSTP-----LRTPLWNGS-NHSTETQSSSEIEIVPSPSP 73  
DB 46 GHPPSGCSTPSPASVQACQRTTGGSOFAASTKWTSPSLNAIETQSTSEELVPSPSP 105  
QY 74 PPLPRIYKPCFVQODKSSGHHYGSACGCKGFFRRSIQKNMYTCHRDKNCIINKYTRN 133  
DB 106 LPPRYIKPCFVQODKSSGHHYGSACGCKGFFRRSIQKNMYTCHRDKNVINKYTRN 165  
QY 134 PCOYCRLOKCFEYGMKSEYRNRNKKKVPKPCSESYTLTPEVGELTEKVRKAHOET 193  
DB 166 RCOYCRLOKCFEYGMKSEYRNRNKKKVPKPCSESYTLTPEVGELTEKVRKAHOET 225  
QY 194 FPALCQLGKYTTNNSRQYSLDLMDKFSLSKCIKIVFAKQLPGFTTLTIADQI 253  
DB 226 FPSLCQLGKYTTNNSRQYSLDLMDKFSLSKCIKIVFAKQLPGFTTLTIADQI 285  
QY 254 TLKAAACLDLILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFPLDLVAFANQLLP 313  
DB 286 TLKAAACLDLILIRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFPLDLVAFANQLLP 345  
QY 314 LEMDDAETGLLSAICLCGRQDLQEPDRVDMLOEPLEALKVYVRKRRSPRHPFKML 373  
DB 346 LEMDDAETGLLSAICLCGRQDLQEPDRVDMLOEPLEALKVYVRKRRSPRHPFKML 405  
QY 374 MKITDLRSISAKGAERVITLTKMEIPGSMPLIQEMLNSENGLDLSGPGGGGGRGLP 433  
DB 406 MKITDLRSISAKGAERVITLTKMEIPGSMPLIQEMLNSENGLDLSGPGGGGGRGLP 461  
QY 434 PPPGSCPSLSPSNRSPATHSP 457  
DB 462 -----SPSVSPSSVNGSVSQSP 479

RESULT 9  
ID RRB\_COTJA STANDARD; PRT; 455 AA.  
AC Q9W6B3;  
DT 15-JUN-2002 (Rel. 41, Created)  
DT 15-JUN-2002 (Rel. 41, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE Retinoic acid receptor beta (RAR-beta).  
GN RARB OR NR1B2.  
OS Coturnix coturnix japonica (Japanese quail).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Coturnix.  
OX NCBI\_TaxID=93934;  
RN SEQUENCE FROM N.A.  
RP Fu Z.W., Kato H., Sugahara K., Kubo T.;  
RT "Molecular cloning of two isoforms of Japanese quail RAR beta and  
RT their expression profile during embryogenesis and in the developing  
RT tissues.";  
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE  
CC HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS  
CC A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS  
CC CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.  
CC -1- SUBCELLULAR LOCATION: Nuclear.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; BETA-1 (SHOWN HERE) AND BETA-2;  
CC ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,  
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.  
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.  
CC NR1 SUBFAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).  
CC  
CC EMBL; AF110729; AAD23397.1; -  
DR HSP; P10826; 1HRA.  
DR InterPro; IPR000536; Hormone\_rec\_lig.  
DR InterPro; IPR001628; Znf\_C4steroid.  
DR Pfam; PF00104; hormone\_rec; 1.  
DR Pfam; PF00105; zf-C4; 1.  
DR PRINTS; PR00047; STNROIDFINGER.  
DR PRODOM; PD000035; Znf\_C4steroid; 1.  
DR SMART; SM00430; HOL1; 1.  
DR SMART; SM00399; ZNF\_C4; 1.  
DR PROSITE; PS00031; NUCLEAR\_RECEPTOR; 1.  
KW Receptor; Transcription regulation; DNA-binding; Nuclear protein;  
KW Zinc-finger; Multigene family; Alternative splicing.  
FT DOMAIN 1 87 MODULATING (BY SIMILARITY).  
FT DNA\_BIND 88 153 NUCLEAR RECEPTOR-TYPE.  
FT ZN\_FING 88 108 C4-TYPE.  
FT ZN\_FING 124 148 C4-TYPE.  
FT DOMAIN 154 199 HINGE (BY SIMILARITY).  
FT DOMAIN 200 419 LIGAND-BINDING (BY SIMILARITY).  
SQ SEQUENCE 455 AA; 50694 MW; 608951B106C4D41 CRC64;  
Query Match 71.2%; Score 1736; DB 1; Length 455;  
Best Local Similarity 77.3%; Pred. No. 1.5e-120;  
Matches 331; Conservative 32; Mismatches 41; Indels 24; Gaps 2;  
QY 30 PEKGLPAPGYPSTPLRTPLWNGSNHSIETQSSSEIEIVPSPSPPLPRIYKPCFVQODK 89  
DB 49 PTCGCTSP-----ASVETQSTSEELVPSPSPPLPPRYKPCFVQODK 94  
QY 90 SSGYHYGSACGCKGFFRRSIQKNMYTCHRDKNCIINKYTRNPOCYCRLOKCFEYGM 149  
DB 95 SSGYHYGSACGCKGFFRRSIQKNMYTCHRDKNCIINKYTRNPOCYCRLOKCFEYGM 154  
QY 150 KESVNRNKKKVPKPCSESYTLTPVGVGLIEKVRKAHOETFPALCQLGKYTTNNS 209  
DB 155 KESVNRNKKKVPKPCSESYTLTPVGVGLIEKVRKAHOETFPALCQLGKYTTNNS 214  
QY 210 EQRSLDLMDKFSLSKCIKIVFAKQLPGFTTLTIADQITLLKAAACLDLILIRIC 269  
DB 215 DHRVLDLMDKFSLSKCIKIVFAKQLPGFTTLTIADQITLLKAAACLDLILIRIC 274  
QY 270 TRYTPEDQMTFSDGLTLNRTQMHNAGFPLDLVAFANQLPLEMDDETGLLSAICL 329  
DB 275 TRYTPEDQMTFSDGLTLNRTQMHNAGFPLDLVAFANQLPLEMDDETGLLSAICL 334  
QY 330 ICGDRQDLQEPDRVDMLOEPLEALKVYVRKRRSPRHPFKMLKITDLRSISAKGAER 389  
DB 335 ICGDRQDLQEPDRVDMLOEPLEALKVYVRKRRSPRHPFKMLKITDLRSISAKGAER 394  
QY 390 VITLTKMEIPGSMPLIQEMLNSENGLDLSGPGGGGGRGLPPLPPGSCPSLSPSNR 449  
DB 395 VITLTKMEIPGSMPLIQEMLNSENGLDLSGPGGGGGRGLPPLPPGSCPSLSPSNR 444  
QY 450 SSPATHSP 457  
DB 445 NSSVSQSP 452  
RESULT 10  
ID RRB\_CHICK STANDARD; PRT; 455 AA.  
AC P22448; Q90598; P27537; Q91354;  
DT 01-AUG-1991 (Rel. 19, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)  
DE Retinoic acid receptor beta (RAR-beta).  
GN RARB OR NR1B2.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;









```
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M24857; ; NOT_ANNOTATED_CDS.
CC EMBL; M32074; AAA60253.1; ;
CC PDB; 3LBD; 02-MAR-99.
CC PDB; 4LBD; 02-MAR-99.
CC TRANSFAC; T00720; ;
CC Genew; HGNC:9866; RARG.
CC MIM; 180190; ;
CC InterPro; IPR000536; Hormone_rec_lig.
CC InterPro; IPR001723; Stdhrmn_receptor.
CC InterPro; IPR001628; Znf_C4steroid.
CC Pfam; PF00104; hormone_rec; 1.
CC Pfam; PF00105; zf-C4; 1.
CC PRINTS; PR00398; STRDHORMONER.
CC PRINTS; PR00047; STROIDFINGER.
CC PRODOM; PD000035; Znf_C4steroid; 1.
CC SMART; SM00430; HOLI; 1.
CC SMART; SM00399; Znf_C4; 1.
CC PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
CC Receptor; Transcription regulation; DNA-binding; Nuclear protein;
CC Zinc-finger; Multigene family; Alternative splicing; 3D-structure.
CC DOMAIN 1 78
CC FT DNA_BIND 79 144 NUCLEAR RECEPTOR-TYPE.
CC FT ZN_FING 79 99 C4-TYPE.
CC FT ZN_FING 115 139 C4-TYPE.
CC FT DOMAIN 145 190 HINGE.
CC FT DOMAIN 191 410 LIGAND-BINDING.
CC SEQUENCE 443 AA; 49307 MW; 4D709194F5111E86 CRC64;
Query Match 67.9%; Score 1655.5; DB 1; Length 443;
Best Local Similarity 71.6%; Pred. NO. 1.2e-114;
Matches 323; Conservative 35; Mismatches 68; Indels 25; Gaps 4;
Qy 1 MYESVEVGPTNPPLVVDYFNQNRACLLPEKGL---PAPGYSYPLRPLW--NGSNHS 55
Db 1 MYDCHETAPGPRLY-----GAGPGAGLLRRATGSCFAGLESFAPQPASLOS 51
Qy 56 IETQSSSEIEVPSPPPPLPRYKPCFVCQDKSSGYHYGSACEGCKGFFRRSIQKNM 115
Db 52 VETQSTSEEMVSPSPPPPPRVYKPCFVCNDXSSGYHYGSCEGCKGFFRRSIQKNM 111
Qy 116 VYTCRDKNCINKYTRNQCQCRLOKCFEVMGSKESVRNDRNKKKVKPECSSEYTL 175
Db 112 VYTCRDKNCINKYTRNQCQCRLOKCFEVMGSKESVRNDRNKKKVKPECSSEYTL 171
Qy 176 TPEVCELTEKYKAHOETFPALCOLGKYTTNNSQPSQVSLDLDLWCKSELTKIITV 235
Db 172 SPQLEELTKYKAHOETFPALCOLGKYTTNNSQPSQVSLDLDLWCKSELTKIITV 231
Qy 236 EFAKQLPGFTTLTADQITLLKACLDLILRICTRYPTPEQDTMTFSGLTLNRTQMNA 295
Db 232 EFAKRLPGFTGLSTADQITLLKACLDLILRICTRYPTPEQDTMTFSGLTLNRTQMNA 291
Qy 296 FGFLPTDLVFAFANQLPLEMDADTGLLSAICLCGDRQDLQEPDRVMDLQEPLEALK 355
Db 292 FGFLPTDLVFAFAGQLPLEMDADTGLLSAICLCGDRMDLEEKVQKQEPLEALR 351
Qy 356 VYVRRKRSRPHMPKMLKKTDLRSISAKGERVITLKMELPGSMPLIQEMLNSEL 415
Db 352 LYARRRPSQYMPFRLMKITDLRGISTKGAERAITLKMELPGMPPLIREMLNPEMF 411
Qy 416 DTLSGQPG-----GGRRGGGLPPP 435
Db 412 EDDSSQCPGPHPNASSEDEVPQGGKGLKSP 442
```

```
RESULT 14
RGL_MOUSE STANDARD; PRT; 458 AA.
ID RGL_MOUSE AC P18911;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Retinoic acid receptor gamma-A (RAR-gamma-A).
GN RARG OR NR1B3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89295563; PubMed=2544807;
RA Zelent A., Krust A., Petkovich M., Kastner P., Chambon P.;
RT "Cloning of murine alpha and beta retinoic acid receptors and a novel
RT receptor gamma predominantly expressed in skin.";
RL Nature 339:714-717(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90220619; PubMed=2157970;
RA Giguere V., Shago M., Zirngibl R., Tate P., Rossant J., Varmuza S.;
RT "Identification of a novel isoform of the retinoic acid receptor
RT gamma expressed in the mouse embryo.";
RL Mol. Cell. Biol. 10:2335-2340(1990).
RN [3]
RP SEQUENCE OF 1-74 FROM N.A.
RX MEDLINE=90207264; PubMed=2157210;
RA Kastner P., Krust A., Mendelsohn C., Garnier J.M., Zelent A.,
RA Leroy P., Staub A., Chambon P.;
RT "Murine isoforms of retinoic acid receptor gamma with specific
RT patterns of expression.";
RL Proc. Natl. Acad. Sci. U.S.A. 87:2700-2704(1990).
CC -!- FUNCTION: THIS IS A RECEPTOR FOR RETINOIC ACID. THIS METABOLITE
CC HAS PROFOUND EFFECTS ON VERTEBRATE DEVELOPMENT. RETINOIC ACID IS
CC A MORPHOGEN AND IS A POWERFUL TERATOGEN. THIS RECEPTOR CONTROLS
CC CELLS FUNCTIONS BY DIRECTLY REGULATING GENE EXPRESSION.
CC -!- SUBCELLULAR LOCATION: Nuclear.
CC -!- ALTERNATIVE PRODUCTS: 2 ISOFORMS; RAR-GAMMA-A (SHOWN HERE) AND
CC RAR-GAMMA-B (AC P20787); ARE PRODUCED BY ALTERNATIVE SPLICING.
CC -!- DOMAIN: COMPOSED OF THREE DOMAINS: A MODULATING N-TERMINAL DOMAIN,
CC A DNA-BINDING DOMAIN AND A C-TERMINAL STEROID-BINDING DOMAIN.
CC -!- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTOR FAMILY.
CC NR1 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X15848; CAA33845.1; ;
CC EMBL; M34476; AAA40035.1; ;
CC EMBL; M32068; AAA40032.1; ;
CC PIR; S05052; S05052.
CC PIR; A34714; A34714.
CC PIR; A35991; A35991.
CC HSSP; P22932; 3LBD.
CC TRANSFAC; T00717; ;
CC TRANSFAC; T01340; ;
CC MGD; MGI:97858; Rarg.
CC InterPro; IPR000536; Hormone_rec_lig.
CC InterPro; IPR001723; Stdhrmn_receptor.
CC InterPro; IPR001628; Znf_C4steroid.
CC Pfam; PF00104; hormone_rec; 1.
CC Pfam; PF00105; zf-C4; 1.
CC PRINTS; PR00398; STRDHORMONER.
CC PRINTS; PR00047; STROIDFINGER.
```



```
Db 228 IKIVEFAKRLFGFTGLSIADQITLLKAACLDILMLRICTRYTPQDWTFTSDGLTLNRQ 287
Qy 292 MHNAGFGPLTDLVFAFANOLLPLEMDDAETGLLSAICLICGDRQDLEQPDVDMLOEPLL 351
Db 288 MHNAGFGPLTDLVFAFAGQLLPLEMDTETGLLSAICLICGDRMDLEEPEKYDKLOEPLL 347
Qy 352 EALKVYVRRRPSRPHMFKMLKITDLSISAKGAERVITLKMEIPGSMPLIOEMLEN 411
Db 348 EALRLYARRDPAPKPYMFPRLMKITDLRGISTKGAERAITLKMEIPGMPPLIREMLN 407
Qy 412 SEGDLTSGQPG-----GGGRDGGGLPPPP 436
Db 408 PEMFEDDSKFGPHPKASSEDEAPGGQKRGQSPQ 443
```

Search completed: March 29, 2003, 06:39:39  
Job time : 24 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2003, 06:37:53 ; Search time 35 seconds  
(without alignments)  
1255.241 Million cell updates/sec

Title: US-09-691-220-2

Perfect score: 2437

Sequence: 1 MYSEVEGGTPNPFLVDF.....SCPSLSLSPSSNRSSPAPHSP 457

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR\_73:\*

1: piri:\*

2: piri2:\*

3: piri3:\*

4: piri4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2127	87.3	452	1 A29491	retinoic acid rece
2	2124.5	87.2	452	2 S78481	retinoic acid rece
3	2119	87.0	955	4 C40045	probable transcrip
4	2099	86.1	453	2 I50574	retinoic acid rece
5	2089	85.7	462	2 S05050	retinoic acid rece
6	2058.5	84.5	464	2 A56558	retinoic acid rece
7	2026.5	83.2	448	2 B56558	retinoic acid rece
8	2008.5	82.4	444	2 I51256	retinoic acid rece
9	1972	80.9	458	2 S06123	retinoic acid rece
10	1926.5	79.1	459	2 A41977	retinoic acid rece
11	1779	73.0	448	2 S02827	retinoic acid rece
12	1771	72.7	448	2 S05051	retinoic acid rece
13	1757	72.1	448	2 A43786	retinoic acid rece
14	1733	71.1	455	2 S13512	retinoic acid rece
15	1663.5	68.3	499	2 I51257	retinoic acid rece
16	1662	68.2	454	1 A33903	retinoic acid rece
17	1655.5	67.9	443	1 C35991	retinoic acid rece
18	1651	67.7	458	2 A34714	retinoic acid rece
19	1633.5	67.0	447	2 B34714	retinoic acid rece
20	1627.5	66.8	442	2 A38592	retinoic acid rece
21	1621	66.5	454	2 S06124	retinoic acid rece
22	1620.5	66.5	476	2 B41977	retinoic acid rece
23	724	29.7	158	2 S02758	retinoic acid rece
24	658	27.0	410	2 S09178	retinoic acid rece
25	657.5	27.0	418	1 TVXLTA	thyroid hormone re
26	653	26.8	410	2 S14416	thyroid hormone re
27	653	26.8	418	2 B36067	thyroid hormone re
28	652	26.8	410	1 QRWSA1	thyroid hormone re
29	652	26.8	410	2 A40917	thyroid hormone re

30	646.5	26.5	391	2	I51097	thyroid hormone re
31	646	26.5	416	2	I51096	thyroid hormone re
32	645.5	26.5	408	1	TVCHVR	thyroid hormone re
33	639	26.2	373	2	I51165	gene c-erba-beta p
34	638	26.2	410	2	S06410	thyroid hormone re
35	630	25.9	373	2	C36067	thyroid hormone re
36	625.5	25.7	475	2	A57035	thyroid hormone be
37	625	25.6	373	2	D36067	thyroid hormone re
38	619.5	25.4	456	1	TVHUAR	thyroid hormone re
39	618	25.4	369	1	TVCHTB	thyroid hormone re
40	618	25.4	369	2	S58211	beta-thyroid hormo
41	615	25.2	461	2	A40377	thyroid hormone re
42	615	25.2	461	2	A31820	thyroid hormone re
43	614.5	25.2	398	1	TVFVVR	transforming prote
44	612.5	25.1	406	2	A44776	transforming prote
45	593.5	24.4	413	2	S14418	thyroid hormone re

ALIGNMENTS

RESULT 1

A29491

retinoic acid receptor alpha - human

C:Species: Homo sapiens (man)

C:Date: 04-Jun-1999 #sequence\_revision 04-Jun-1999 #text\_change 16-Jun-2000

C:Accession: A29491; S07272; S15598; A58975; A29740; S15599

R:Giguere, V.; Ong, E.S.; Segui, P.; Evans, R.M.

Nature 330, 624-629, 1987

A:Title: Identification of a receptor for the morphogen retinoic acid.

A:Reference number: A29491; MUID:88065922; PMID:2825036

A:Accession: A29491

A:Molecule type: mRNA

A:Residues: 1-462 <GIG>

A:Cross-references: GB:X06614; NID:g36156; PIDN:CAA29829.1; PID:g36157

R:Chambon, P. A human retinoic acid receptor which belongs to the family of nuclear recept

submitted to the EMBL Data Library, December 1988

A:Reference number: S07272

A:Accession: S07272

A:Molecule type: mRNA

A:Residues: 31-462 <CHA>

A:Cross-references: EMBL:X06538; NID:g35873; PIDN:CAA29787.1; PID:g35874

R:Petkovich, M.; Brand, N.J.; Krust, A.; Chambon, P.

Nature 330, 444-450, 1987

A:Title: A human retinoic acid receptor which belongs to the family of nuclear recept

A:Reference number: S06224; MUID:88065872; PMID:2825025

A:Accession: S06224

A:Molecule type: mRNA

A:Residues: 31-179, 'V', 181-462 <PET>

A:Cross-references: EMBL:X06538

A:Note: This sequence has been revised in reference S07272

A:Note: the authors translated the codon AAC for residue 360 as Lys; the sequence sho

R:Brand, N.J.; Petkovich, M.; Chambon, P.

Nucleic Acids Res. 18, 6799-6806, 1990

A:Title: Characterization of a functional promoter for the human retinoic acid recept

A:Reference number: S15594; MUID:91088249; PMID:2175878

A:Accession: S15598

A:Molecule type: DNA

A:Residues: 1-80 <BRAL>

A:Cross-references: EMBL:X56058; NID:g35876; PIDN:CAA39533.1; PID:g825712

A:Accession: A58975

A:Molecule type: DNA

A:Residues: 61-80 <BRA2>

A:Cross-references: GB:X58685; NID:g35878; PIDN:CAA41532.1; PID:gl335286

C:Genetics:

A:Gene: GDB:RARA

A:Cross-references: GDB:I20337; OMIM:I80240

A:Map position: 17q12-17q12

A:Introns: 60/1

A:Note: the first intron occurs before the initiator codon; the list of introns is in

C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology.

C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger

F:86-337/Domain: erba transforming protein homology <ERBA>

F:88-108/Region: zinc finger  
F:124-148/Region: zinc finger

Query Match 87.3%; Score 2127; DB 1; Length 462;  
Best Local Similarity 91.6%; Pred. No. 1.2e-148;  
Matches 413; Conservative 5; Mismatches 13; Indels 20; Gaps 4;

QY 7 VGGTPNPFLVDFYNONRACLPEKGLPAGPYSTPLRTPLWNGSNHSIETQSSSBEI 66  
DB 32 LGLLSP-PGALTTLHQ-----LPVSG-YSTP-----SPATIETQSSSBEI 71

QY 67 VPSPSPPLPRIYKPCFCVQDKSSGYHYGSACEGCKGFFRRSIQKNMYTCHRDNCI 126  
DB 72 VPSPSPPLPRIYKPCFCVQDKSSGYHYGSACEGCKGFFRRSIQKNMYTCHRDNCI 131

QY 127 INKVTNRCQYRLQKCFEYVMSKESVYRNDNRNKKKKVPEKPECSYTLTPEVGELETKV 186  
DB 132 INKVTNRCQYRLQKCFEYVMSKESVYRNDNRNKKKKVPEKPECSYTLTPEVGELETKV 191

QY 187 RKAQETFPALCOLGKYTTNNSSEQRVSLDLDLWDFSELSTKCIITVEPAKOLPGFTT 246  
DB 192 RKAQETFPALCOLGKYTTNNSSEQRVSLDLDLWDFSELSTKCIITVEPAKOLPGFTT 251

QY 247 LTADQITLLKAACLDLILIRICTRYTPEQDTMTFSDGLTLNRTOMHNAFGP 306  
DB 252 LTADQITLLKAACLDLILIRICTRYTPEQDTMTFSDGLTLNRTOMHNAFGP 311

QY 307 FANOLLPLEMDDAETGLLSAICLCGRDQLEQPDVDMLOEPLLEALKVYVRKRRSRP 366  
DB 312 FANOLLPLEMDDAETGLLSAICLCGRDQLEQPDVDMLOEPLLEALKVYVRKRRSRP 371

QY 367 HMFPPKMLKITDLRSISAKGAERVITLKMETPGSMPLIQEMLNSEGLDITLSCQPGGG 426  
DB 372 HMFPPKMLKITDLRSISAKGAERVITLKMETPGSMPLIQEMLNSEGLDITLSCQPGGG 431

QY 427 RDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
DB 432 RDGGGLAPPGSCPSLSPSSNRSSPATHSP 462

RESULT 2  
S78481  
retinoic acid receptor alpha-2 - eastern newt  
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)  
C:Date: 11-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 20-Aug-1999  
C:Accession: S78481; S26656  
R:Ragsdale, C.W.  
submitted to the EMBL Data Library, August 1992  
A:Reference number: S78481  
A:Accession: S78481  
A:Molecule type: mRNA  
A:Residues: 1-452 <RAG>  
A:Cross-references: EMBL:Z14254; NID:964123; PIDN:CAA78621.1; PID:964124  
R:Ragsdale Jr., C.W.; Gates, P.B.; Brookes, J.P.  
Nucleic Acids Res. 20, 5851, 1992  
A:Title: Identification and expression pattern of a second isoform of the newt alpha ret  
A:Reference number: S26656; MUID:93087213; PMID:1333599  
A:Accession: S26656  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-52 <RAW>  
A:Cross-references: EMBL:Z14254  
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology  
C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger  
F:79-331/Domain: erba transforming protein homology <ERBA>  
F:81-101/Region: zinc finger CCCC motif  
F:117-141/Region: zinc finger CCCC motif

Query Match 87.2%; Score 2124.5; DB 2; Length 452;  
Best Local Similarity 87.1%; Pred. No. 1.8e-148;  
Matches 399; Conservative 29; Mismatches 23; Indels 7; Gaps 3;  
QY 1 MYESVEVGTPNPFLVDFYNONRACLPEKGLPAGPYSTPLRTPLWNGSNHSIETQS 60

DB 1 MYDSVEVS--SPSPYIMIDFYSONRACLMDKGLGHPVPFGSPINRPHWSSSHSIETQS 58  
QY 61 SSSSEIVSPSPPLPRIYKPCFCVQDKSSGYHYGSACEGCKGFFRRSIQKNMYTCH 120  
DB 59 TSSEIVSPSPPLPRIYKPCFCVQDKSSGYHYGSACEGCKGFFRRSIQKNMYTCH 118  
QY 121 RDKNCIINKVTNRCQYRLQKCFEYVMSKESVYRNDNRNKKKKVPEKPECSYTLTPEV 179  
DB 119 RDKTCIINKVTNRCQYRLQKCFEYVMSKESVYRNDNRNKKKKVPEKPECSYTLTPEV 178  
QY 180 GELTEKVRKAHQETFPALCOLGKYTTNNSSEQRVSLDLDLWDFSELSTKCIITVEPAK 239  
DB 179 EDLVEKVRKAHQETFPALCOLGKYTTNNSSEQRVSLDLDLWDFSELSTKCIITVEPAK 238  
QY 240 QLPGFTTITADQITLLKAACLDLILIRICTRYTPEQDTMTFSDGLTLNRTOMHNAFGP 299  
DB 239 QLPGFTTITADQITLLKAACLDLILIRICTRYTPEQDTMTFSDGLTLNRTOMHNAFGP 298  
QY 300 LTDLVFAFANOLLPLEMDDAETGLLSAICLCGRDQLEQPDVDMLOEPLLEALKVYVR 359  
DB 299 LTDLVFAFANOLLPLEMDDAETGLLSAICLCGRDQLEQPDVDMLOEPLLEALKVYVR 358  
QY 360 KRRPSRPHMFPKMLKITDLRSISAKGAERVITLKMETPGSMPLIQEMLNSEGLDITLS 419  
DB 359 KRRPNKPHMFPKMLKITDLRSISAKGAERVITLKMETPGSMPLIQEMLNSEGLDITLS 418  
QY 420 GQPGGGRRDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
DB 419 GQP-----PRASSLAPPFGSCPSLSPSSNRSSPATHSP 452

RESULT 3  
C40045  
probable transcription factor PML/retinoic acid receptor alpha mutant fusion protein  
C:Species: Homo sapiens (man)  
C:Date: 05-Jun-1992 #sequence\_revision 17-Nov-1995 #text\_change 20-Apr-2000  
C:Accession: C40045; B40044  
R:de The, H.; Lavau, C.; Marchio, A.; Chomienne, C.; Degos, L.; Dejean, A.  
Cell 66, 675-684, 1991  
A:Title: The PML-RARalpha fusion mRNA generated by the t(15;17) translocation in acu  
A:Reference number: A40045; MUID:91347369; PMID:1652369  
A:Accession: C40045  
A:Molecule type: mRNA  
A:Residues: 1-955 <DET>  
A:Cross-references: GB:S50916; NID:Q234248; PIDN:AAB19602.1; PID:Q234249  
R:Kakizuka, A.; Miller Jr., W.H.; Umesono, K.; Warrell Jr., R.P.; Frankel, S.R.; Muri  
Cell 66, 663-674, 1991  
A:Title: Chromosomal translocation t(15;17) in human acute promyelocytic leukemia fu  
A:Reference number: A40044; MUID:91347368; PMID:1652368  
A:Accession: B40044  
A:Molecule type: mRNA  
A:Residues: 1-394,553-955 <KAK>  
A:Cross-references: GB:M73779; NID:gl90124; PIDN:AAA60126.1; PID:gl90125  
C:Comment: This sequence is the chimeric product of a translocation mutation.  
C:Genetics:  
A:Gene: PML/RAR  
A:Map position: 15q22/17q12  
C:Keywords: alternative splicing; fusion protein  
F:57-227/Region: cysteine-rich  
F:581-601/Region: zinc finger motif  
F:617-641/Region: zinc finger motif

Query Match 87.0%; Score 2119; DB 4; Length 955;  
Best Local Similarity 90.9%; Pred. No. 1.2e-147;  
Matches 409; Conservative 7; Mismatches 12; Indels 22; Gaps 3;  
QY 30 PEKGLPA-----PGPYSTPLRTPLW-NGSNH-----SIETQSSSSEIV 67  
DB 506 PEQPRPSTSKAVSPHPLDGPSPRSPVIGSEVFLPNSNHNHVASGAGAAIETQSSSEIV 565  
QY 68 PSPSPSPPLPRIYKPCFCVQDKSSGYHYGSACEGCKGFFRRSIQKNMYTCHRDNCII 127

Db 566 PSPSPPLPRIYKPCFVCODKSSGYHYGSACEGCKGFFRRSIQKNMVTCHRDKNCI 625  
QY 128 NKVTRNQCQYCRLOKCFEYVGSKESVNRDNKKKEVPKPCSESYTTPVEGELIEKYR 187  
Db 626 NKVTRNQCQYCRLOKCFEYVGSKESVNRDNKKKEVPKPCSESYTTPVEGELIEKYR 685  
QY 188 KAHQETPALCOLGKYTTNNSEORVSLDIDLWDFSELSKCIITKTFEAKQLPGFTTL 247  
Db 686 KAHQETPALCOLGKYTTNNSEORVSLDIDLWDFSELSKCIITKTFEAKQLPGFTTL 745  
QY 248 TIADQITLLKAACDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFGPLTDLVFAF 307  
Db 746 TIADQITLLKAACDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFGPLTDLVFAF 805  
QY 308 ANQLLPLEMDDAETGLLSAICLICGDRDLEOPDRVDMQLQPLLBALKVYVYKRPRSRPH 367  
Db 806 ANQLLPLEMDDAETGLLSAICLICGDRDLEOPDRVDMQLQPLLBALKVYVYKRPRSRPH 865  
QY 368 MFPKMLMKITDLRSISAKGAERVITLKEIPGSMPLIQEMLENSEGLDLSGQPGGGGR 427  
Db 866 MFPKMLMKITDLRSISAKGAERVITLKEIPGSMPLIQEMLENSEGLDLSGQPGGGGR 925  
QY 428 DGGGLPPPGSCSPSLSPSSNRSSPATHSP 457  
Db 926 DGGGLAPPGSCSPSLSPSSNRSSPATHSP 955  
RESULT 4  
I50674  
retinoic acid receptor alpha isoform 2-1 - chicken  
C:Species: Gallus gallus (chicken)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 20-Aug-1999  
C:Accession: I50674  
R:Michaille, J.J.; Kanzler, B.; Blanchet, S.; Garnier, J.M.; Dhoulailly, D.  
Int. J. Dev. Biol. 39, 587-596, 1995  
A:Title: Characterization of cDNAs encoding two chick retinoic acid receptor alpha isoforms  
A:Reference number: I50674; MUID:96192630; PMID:8619957  
A:Accession: I50674  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-453 <MIC>  
A:Cross-references: EMBL:X78335; NID:q1089845; PIDN:CAA55134.1; PID:q1089846  
R:Heiermann, R.; Rentrop, M.; Lang, E.; Maelicke, A.  
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology  
C:Keywords: zinc finger  
F:79-330/Domain: erba transforming protein homology <ERBA>  
Query Match 86.1%; Score 2099; DB 2; Length 453;  
Best Local Similarity 88.0%; Pred. No. 1.4e-146; Mismatches 36; Indels 6; Gaps 4;  
Matches 403; Conservative  
QY 1 MYSEVEGG-PTNPFLVDFYNONRACLLPEKGLPAPGPYSTPLRTPLWNGSNHSIETQ 59  
Db 1 MFEAEVAGLPPPGPLPRMDCGCGRGLLPQ--CPPP-PTAPRAPHPWASGRSVETQ 57  
QY 60 SSSSEIVPSPSPPLPRIYKPCFVCODKSSGYHYGSACEGCKGFFRRSIQKNMVTTC 119  
Db 58 STSSEIVPSPSPPLPRIYKPCFVCODKSSGYHYGSACEGCKGFFRRSIQKNMVTTC 117  
QY 120 HRDKNCLINKVTRNQCQYCRLOKCFEYVGSKESVNRDNKKKEVPKPCSESYTTPVEV 179  
Db 118 HRDKNCLINKVTRNQCQYCRLOKCFEYVGSKESVNRDNKKKEVPKPCSESYTTPVEV 177  
QY 180 GELLEKVKHAHOEFPAFCALCOLGKYTTNNSEORVSLDIDLWDFSELSKCIITKTFEAK 239  
Db 178 EELLEKVKHAHOEFPAFCALCOLGKYTTNNSEORVSLDIDLWDFSELSKCIITKTFEAK 237  
QY 240 QLPGFTTLTADQITLLKAACDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFGP 299  
Db 238 QLPGFTTLTADQITLLKAACDLILIRICTRYTPQDTMTFSDGLTLNRTQMHNAGFGP 297  
QY 300 LTDLVFAFANQLLPLEMDDAETGLLSAICLICGDRDLEOPDRVDMQLQPLLBALKVYVYKR 359  
Db 300 LTDLVFAFANQLLPLEMDDAETGLLSAICLICGDRDLEOPDRVDMQLQPLLBALKVYVYKR 359

Db 298 LTDLVFAFANQLLPLEMDDAETGLLSAICLICGDRDLEOPDRVDMQLQPLLBALKVYVYKR 357  
QY 360 KRRPRSRPHMFKMLMKITDLRSISAKGAERVITLKEIPGSMPLIQEMLENSEGLDLS 419  
Db 358 KRRPRSRPHMFKMLMKITDLRSISAKGAERVITLKEIPGSMPLIQEMLENSEGLDLS 417  
QY 420 GPGGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457  
Db 418 GPGGPRTP--GGGLGPPPGSCSPSLSPSSNRSSPATHSP 453  
RESULT 5  
S05050  
retinoic acid nuclear receptor isoform alpha 1 - mouse  
C:Species: Mus musculus (house mouse)  
C:Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 20-Aug-1999  
C:Accession: S05050; S23797; S13599; I56594  
R:Zelent, A.; Krust, A.; Petkovich, M.; Kastner, P.; Chambon, P.  
Nature 339, 714-717, 1989  
A:Title: Cloning of murine alpha and beta retinoic acid receptors and a novel receptor  
A:Reference number: S05050; MUID:89295363; PMID:2544807  
A:Accession: S05050  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-462 <ZEL>  
R:Kruyt, F.A.E.; van der Veer, L.J.; Mader, S.; van den Brink, C.E.; Feiljen, A.; Jor  
Differentiation 49, 27-37, 1992  
A:Title: Retinoic acid resistance of the variant embryonal carcinoma cell line RAC65  
A:Reference number: S23797; MUID:92324516; PMID:1320576  
A:Accession: S23797  
A:Molecule type: mRNA  
A:Residues: 1-390, 'A' <KRU>  
A:Cross-references: EMBL:X57528; NID:g53890; PIDN:CAA04749.1; PID:g53891  
A:Experimental source: embryonic carcinoma cell line RAC65  
A:Note: this is the entire sequence of a truncated retinoic acid receptor alpha proc  
R:Leroy, P.; Krust, A.; Zelent, A.; Mendelsohn, C.; Garnier, J.M.; Kastner, P.; Dier  
EMBO J. 10, 59-69, 1991  
A:Title: Multiple isoforms of the mouse retinoic acid receptor alpha are generated t  
A:Reference number: S13599; MUID:91114713; PMID:1846598  
A:Accession: S13599  
A:Molecule type: mRNA  
A:Residues: 1-73 <LER>  
A:Cross-references: EMBL:X56572; NID:g53892; PIDN:CAA39919.1; PID:g53893  
R:Heiermann, R.; Rentrop, M.; Lang, E.; Maelicke, A.  
J. Recept. Res. 13, 693-709, 1993  
A:Title: Cloning of several genes coding for retinoic acid nuclear receptors in the  
A:Reference number: I56594; MUID:93195858; PMID:8383767  
A:Accession: I56594  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-462 <RES>  
A:Cross-references: GB:S56656; NID:q298621; PIDN:AAB25783.1; PID:q298622  
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology  
C:Keywords: alternative splicing; DNA binding; nucleus; transcription regulation; zi  
F:86-337/Domain: erba transforming protein homology <ERBA>  
F:88-108/Region: zinc finger  
F:124-148/Region: zinc finger  
Query Match 85.7%; Score 2089; DB 2; Length 462;  
Best Local Similarity 90.0%; Pred. No. 7.7e-146;  
Matches 406; Conservative 8; Mismatches 17; Indels 20; Gaps 4;  
QY 7 VGGPTNPFLVDFYNONRACLLPEKGLPAPGPYSTPLRTPLWNGSNHSIETQSSSEI 66  
Db 32 LGGLSP-FCALYSLQHQ-----LPVSG-YSTP-----SPATETUSSSEI 71  
QY 67 VSPSPSPPLPRIYKPCFVCODKSSGYHYGSACEGCKGFFRRSIQKNMVTCHRDKNCI 126  
Db 72 VSPSPSPPLPRIYKPCFVCODKSSGYHYGSACEGCKGFFRRSIQKNMVTCHRDKNCI 131  
QY 127 INKVTNRCQYCRLOKCFEYVGSKESVNRDNKKKEVPKPCSESYTTPVEGELIEKY 186  
Db 132 INKVTNRCQYCRLOKCFEYVGSKESVNRDNKKKEVPKPCSESYTTPVEGELIEKY 191

QY 187 RKAHOETFPALCOLGKYYTNNSEORVSLDIDLWDKFSLSLTKCIKIVVERAKQLPGTT 246  
Db 192 RKAHOETFPALCOLGKYYTNNSEORVSLDIDLWDKFSLSLTKCIKIVVERAKQLPGTT 251  
QY 247 LTIADQITLLKAACLDLILIRICTRYTPEODTMTFSDGLTLNRTQMHNAGFPLTDLVFA 306  
Db 252 LTIADQITLLKAACLDLILIRICTRYTPEODTMTFSDGLTLNRTQMHNAGFPLTDLVFA 311  
QY 307 FANQLLPLEMDDAETGLLSAICLCGDRQDLEQPDVMDLQEPLEALKVYVRKRRSRP 366  
Db 312 FANQLLPLEMDDAETGLLSAICLCGDRQDLEQPDVMDLQEPLEALKVYVRKRRSRP 371  
QY 367 HMPFKMLMKITDLRSISAKGAERVITLKEIPGSMPPPLIQEMLNSEGLDITLSGPGGG 426  
Db 372 HMPFKMLMKITDLRSISAKGAERVITLKEIPGSMPPPLIQEMLNSEGLDITLSGPGGG 431  
QY 427 RDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
Db 432 RDGGGLAPPGSCPSLSPSSNRSSPATHSP 462

RESULT 6  
A56558  
retinoic acid receptor isoform alpha 2.1 - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 31-Oct-1997  
C:Accession: A56558  
R:Sharpe, C.R.  
Mech. Dev. 39, 81-93, 1992  
A:Title: Two isoforms of retinoic acid receptor alpha expressed during xenopus developme  
A:Reference number: A56558; MUID:93144180; PMID:1336976  
A:Accession: A56558  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-464 <SHA>  
A:Experimental source: embryos  
A:Note: sequence extracted from NCBI backbone (NCBIN:123865, NCBIP:123867)  
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology  
C:Keywords: zinc finger  
F:91-342/Domain: erba transforming protein homology <ERBA>

Query Match 84.5%; Score 2058.5; DB 2; Length 464;  
Best Local Similarity 85.6%; Pred. No. 1.4e-143;  
Matches 393; Conservative 21; Mismatches 40; Indels 5; Gaps 2;

QY 1 MYSEVEVG--GPTNPPLVDFYNQNRACLLPEKGLPAGPYSTPLRPLNGSNHSTET 58  
Db 9 MYENVDPALASSPTRFHMDFYSHNRQCLLQEKIGITIHPIYGTPLRQTHWSSNHSTET 68  
QY 59 QSSSEIEIVPSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRFSIOKNMYIT 118  
Db 69 QSTSEIEIVPSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRFSIOKNMYIT 128  
QY 119 CHRDKNCIINKVTRNRCQYRLQKCFEVMGSKESVRNDRNKKKEVPKPECESEYTLTPE 178  
Db 129 CHRDKNCIINKVTRNRCQYRLQKCFEVMGSKESVRNDRNKKKEVPKPECESEYTLTPE 188  
QY 179 VGLIEKVRKAHOETFPALCOLGKYYTNNSEORVSLDIDLWDKFSLSLTKCIKIVVERA 238  
Db 189 TDLIEKVRKAHOETFPALCOLGKYYTNNSEORVSLDIDLWDKFSLSLTKCIKIVVERA 248  
QY 239 KQLPGFTTLTADQITLLKAACLDLILIRICTRYTPEODTMTFSDGLTLNRTQMHNAGFG 298  
Db 249 KQLPGFTTLTADQITLLKAACLDLILIRICTRYTPEODTMTFSDGLTLNRTQMHNAGFG 308  
QY 299 PLTDLVFAFANQLLPLEMDDAETGLLSAICLCGDRQDLEQPDVMDLQEPLEALKVYVR 358  
Db 309 PLTDLVFAFANQLVPLEMDDAETGLLSAICLCGDRQDLEQPDVMDLQEPLEALKVYVR 368  
QY 359 RKRPRSRPHMPFKMLMKITDLRSISAKGAERVITLKEIPGSMPPPLIQEMLNSEGLDITL 418  
Db 369 RTRRPQKPHMPFKMLMKITDLRSISAKGAERVITLKEIPGSMPPPLIQEMLNSEGLDITL 428

QY 419 SQPGGGRRDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
Db 429 G---GGASSDAPVTVPAGSCPSLSPSSNRSSPATHSP 464

RESULT 7  
B56558  
retinoic acid receptor isoform alpha 2.2 - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C:Date: 05-Jan-1996 #sequence\_revision 05-Jan-1996 #text\_change 20-Aug-1999  
C:Accession: B56558  
R:Sharpe, C.R.  
Mech. Dev. 39, 81-93, 1992  
A:Title: Two isoforms of retinoic acid receptor alpha expressed during Xenopus devel  
A:Reference number: A56558; MUID:93144180; PMID:1336976  
A:Accession: B56558  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-448 <SHA>  
A:Cross-references: GB:X87365; GB:S53907; GB:S53908; NID:9836654; PIDN:CAA60791.1; P  
A:Experimental source: embryos  
A:Note: sequence extracted from NCBI backbone (NCBIN:123871, NCBIP:123874)  
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology  
C:Keywords: zinc finger  
F:75-326/Domain: erba transforming protein homology <ERBA>

Query Match 83.2%; Score 2026.5; DB 2; Length 448;  
Best Local Similarity 84.9%; Pred. No. 2.9e-141;  
Matches 388; Conservative 22; Mismatches 38; Indels 9; Gaps 3;

QY 1 MYSEVEVGPTNPPLVDFYNQNRACLLPEKGLPAGPYSTPLRPLNGSNHSTETOS 60  
Db 1 MYENVDV---SPTHYHMDDFYSHNRQCLWPEKRI---NPYGTPLGTQHWSSNHSTETOS 54  
QY 61 SSSEIEIVPSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRFSIOKNMYITCH 120  
Db 55 TSSEIEIVPSPSPPLPRIYKPCFVCQDKSSGYHYGVSAECGCKGFFRFSIOKNMYITCH 114  
QY 121 RDKNCIINKVTRNRCQYRLQKCFEVMGSKESVRNDRNKKKEVPKPECESEYTLTPEVG 180  
Db 115 RDKNCIINKVTRNRCQYRLQKCFEVMGSKESVRNDRNKKKEVPKPECESEYTLTPEVQ 174  
QY 181 ELIEKVRKAHOETFPALCOLGKYYTNNSEORVSLDIDLWDKFSLSLTKCIKIVVERAKO 240  
Db 175 DLIEKVRKAHOETFPALCOLGKYYTNNSEORVSLDIDLWDKFSLSLTKCIKIVVERAKO 234  
QY 241 LPGFTTLTADQITLLKAACLDLILIRICTRYTPEODTMTFSDGLTLNRTQMHNAGFGPL 300  
Db 235 LPGFTTLTADQITLLKAACLDLILIRICTRYTPEODTMTFSDGLTLNRTQMHNAGFGPL 294  
QY 301 TDLVFAFANQLLPLEMDDAETGLLSAICLCGDRQDLEQPDVMDLQEPLEALKVYVRK 360  
Db 295 TDLVFAFANQLLPLEMDDAETGLLSAICLCGDRQDLEQPDVMDLQEPLEALKVYVRN 354  
QY 361 RPRSRPHMPFKMLMKITDLRSISAKGAERVITLKEIPGSMPPPLIQEMLNSEGLDITLSG 420  
Db 355 RPRSRPHMPFKMLMKITDLRSISAKGAERVITLKEIPGSMPPPLIQEMLNSEGLDITLSG 413  
QY 421 QPGGGRRDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
Db 414 --GGASSDPTVTPVAPGSCPSLSPSSNRSSPATHSP 448

RESULT 8  
I51256  
retinoic acid receptor alpha - zebra fish  
C:Species: Brachydanio rerio (zebra fish)  
C:Date: 04-Sep-1997 #sequence\_revision 04-Sep-1997 #text\_change 20-Aug-1999  
C:Accession: I51256  
R:Joore, J.; van der Lans, G.B.; Lanser, P.H.; Vervaaert, J.M.; Zivkovic, D.; Speksni  
Mech. Dev. 46, 137-150, 1994  
A:Title: Effects of retinoic acid on the expression of retinoic acid receptors durin



A:Reference number: I51256; MUID:95001557; PMID:7918098  
A:Accession: I51256  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-444 <DOO>  
A:Cross-references: GB:S74155; NID:g704369; PIDN:AA32276.1; PID:g704370  
C:Genetics:  
A:Gene: zRARalpha  
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology  
F:73-324/Domain: erba transforming protein homology <ERBA>  
  
Query Match 82.4%; Score 2008.5; DB 2; Length 444;  
Best Local Similarity 82.9%; Pred. No. 6e-140;  
Matches 379; Conservative 35; Mismatches 30; Indels 13; Gaps 3;  
  
QY 1 MYSEVEGGTPNPLVDFYNQNRACLLPEKGLPAPGYPSTPLRPLWNGSNHSETQS 60  
DB 1 MYSEVDV-----NPLMDYDYNQSGCLIPDK---MPHPFSSIRHQHWSGNSHSETQS 52  
  
QY 61 SSSSEIVPSPSPPLPRIYKPCFCQDKSSGYHGVGSACEGCKGFFRRSIQKNMVTYCH 120  
DB 53 TSSEIVPSPSPSPPLPRIYKPCFCQDKSSGYHGVGSACEGCKGFFRRSIQKNMVTYCH 112  
  
QY 121 RDKNCIINKVTRNRCQYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPEVG 180  
DB 113 REKNCIINKVTRNRCQYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPEVG 172  
  
QY 181 ELIEKVRKAHQETFPALCOLGKYTTNNSSEQRVSIDIDLWDFSELSTKCIITVEFAKQ 240  
DB 173 QMIDRVKKAHQETFPALCOLGKYTTNNSSEQRVSIDIDLWDFSELSTKCIITVEFAKQ 232  
  
QY 241 LPGFTTLTADQITLLKACADILILRICTRYTPQDTMTFSDGLTLNRTQMHNAGFGL 300  
DB 233 LPGFTTLTADQITLLKACADILILRICTRYTPQDTMTFSDGLTLNRTQMHNAGFGL 292  
  
QY 301 TDLVAFANQLLPEMDAETGLLSAICLCGRDROLEQPDVMDLQEPLEALKYVVRK 360  
DB 293 TDLVAFANQLLPEMDAETGLLSAICLCGRDROLEQADKVDVLOEPLEALKYVVRN 352  
  
QY 361 RRPSPHPFPKMLKITDLRSISAKGAERVITLKEIPGSMPLIOEMLENSEGLDTLSG 420  
DB 353 RRPKHPMFPMKMLKITDLRSISAKGAERVITLKEIPGSMPLIOEMLENSEGLSSG 412  
  
QY 421 QPGGGRDGGGLPPPGSCSPSLSPSSNRRSPATHSP 457  
DB 413 AQGRAS-----ATTPGSCSPSLSPNSAQSPPQTQS 444  
  
RESULT 9  
S06123  
retinoic acid receptor alpha - eastern newt  
C:Species: Notophthalmus viridescens, Triturus viridescens (eastern newt)  
C>Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 20-Aug-1999  
C:Accession: S06123  
R:Ragsdale Jr., C.W.; Petkovich, M.; Gates, P.B.; Chambon, P.; Brockes, J.P.  
Nature 341, 654-657, 1989  
A:Title: Identification of a novel retinoic acid receptor in regenerative tissues of the  
A:Reference number: S06123; MUID:90015190; PMID:2552324  
A:Accession: S06123  
A:Molecule type: mRNA  
A:Residues: 1-458 <RAG>  
A:Cross-references: EMBL:X17585; NID:g64125; PIDN:CAA35602.1; PID:g64126  
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology  
C:Keywords: DNA binding; nucleus; transcription regulation; zinc finger  
F:85-337/Domain: erba transforming protein homology <ERBA>  
F:87-107/Region: zinc finger  
F:123-147/Region: zinc finger  
  
Query Match 80.9%; Score 1972; DB 2; Length 458;  
Best Local Similarity 88.3%; Pred. No. 3e-137;  
Matches 379; Conservative 16; Mismatches 18; Indels 16; Gaps 4;  
  
QY 33 GLPAPGP---YSTPLRPLWNGSNHSETQSSESEIVPSPSPPLPRIYKPCFCVQDK 89

DB 42 GIPHPLVSAVSTP-----SPATIEQTSSSEIVPSPSPPLPRIYKPCFCVQDK 93  
QY 90 SSGYHYGVGSACEGCKGFFRRSIQKNMVTYCHRDKNCIINKVTRNRCQYCRLOKCFEVMG 149  
DB 94 SSGYHYGVGSACEGCKGFFRRSIQKNMVTYCHRDKNCIINKVTRNRCQYCRLOKCFEVMG 153  
QY 150 KESVRNDRNKKKK-EVPPKPCSESYTLTPEVGELIEKVRKAHQETFPALCOLGKYTTNNS 208  
DB 154 KESVRNDRNKKKKQKQAPKQECTESYITPEVEDLVEKVRKAHQETFPALCOLGKYTTNNS 213  
QY 209 SQQRVSIDIDLWDFSELSTKCIITVEFAKOLPGFTTLTADQITLLKACADILILRI 268  
DB 214 SEERYSLDIDLWDFSELSTKCIITVEFAKOLPGFTTLTADQITLLKACADILILRI 273  
QY 269 CRYTPEDQDTMTFSDGLTLNRTQMHNAGFGLTDLVAFANQLLPEMDAETGLLSAIC 328  
DB 274 CRYTPEDQDTMTFSDGLTLNRTQMHNAGFGLTDLVAFANQLLPEMDAETGLLSAIC 333  
QY 329 LICGRDQDLEQPDVMDLQEPLEALKYVVRKRRSPRHPFKMLMKITDLRSISAKGAE 388  
DB 334 LICGRDQDLEQPDVMDLQEPLEALKYVVRKRRSPRHPFKMLMKITDLRSISAKGAE 393  
QY 389 RVITLKEIPGSMPLIOEMLENSEGLDTLSGPGGGGGLPPPGSCSPSLSPSSN 448  
DB 394 RVITLKEIPGSMPLIOEMLENSEGLDTLSGPGGGGGLPPPGSCSPSLSPSSN 449  
QY 449 RSPATHSP 457  
DB 450 RSPATHSP 458  
  
RESULT 10  
A41977  
retinoic acid receptor alpha - African clawed frog  
C:Species: Xenopus laevis (African clawed frog)  
C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 07-Feb-1997  
C:Accession: A41977  
R:Blumberg, B.; Mangelsdorf, D.J.; Dyck, J.A.; Bittner, D.A.; Evans, R.M.; De Robertis  
Proc. Natl. Acad. Sci. U.S.A. 89, 2321-2325, 1992  
A:Title: Multiple retinoid-responsive receptors in a single cell: families of retinoid  
A:Reference number: A41977; MUID:92196110; PMID:1312717  
A:Accession: A41977  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 1-459 <BLU>  
A:Experimental source: egg  
A>Note: sequence extracted from NCBI backbone (NCBI:88334)  
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology  
C:Keywords: zinc finger  
F:86-337/Domain: erba transforming protein homology <ERBA>  
F:88-108/Region: zinc finger  
F:124-148/Region: zinc finger  
  
Query Match 79.1%; Score 1926.5; DB 2; Length 459;  
Best Local Similarity 81.6%; Pred. No. 6.7e-134;  
Matches 377; Conservative 17; Mismatches 43; Indels 25; Gaps 4;  
  
QY 10 PTPNPLVDFYNQNRACLLPEK-----GLPAP-----GPYSTPLRPLWNGSNHNS 55  
DB 9 PTPNPLVDFYNQNRACLLPEK-----GLPAP-----GPYSTPLRPLWNGSNHNS 55  
QY 56 IETQSSSEIVPSPSPPLPRIYKPCFCQDKSSGYHYGVGSACEGCKGFFRRSIQKNM 115  
DB 61 IETQSSSEIVPSPSPPLPRIYKPCFCQDKSSGYHYGVGSACEGCKGFFRRSIQKNM 120  
QY 116 VYCHRDKNCIINKVTRNRCQYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTL 175  
DB 121 VYCHRDKNCIINKVTRNRCQYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTL 180  
QY 176 TPEVGELIEKVRKAHQETFPALCOLGKYTTNNSSEQRVSIDIDLWDFSELSTKCIITKV 235  
DB 181 SPETQDILIEKVRKAHQETFPALCOLGKYTTNNSSEQRVSIDIDLWDFSELSTKCIITKV 240



Db 1 MFDCMDVLSVSPGQ--ILDEVTASPSQMLQEKALKACLSGFTQAE--WQHRHTAQSI 55  
QY 58 TQSSSEIIVSPSPPLPRIVKPCFCQDKSSGYHYGVSAACGCGKGFRRSIQKNMY 117  
Db 56 TQSTSSSEIIVSPSPPLPRIVKPCFCQDKSSGYHYGVSAACGCGKGFRRSIQKNMY 115  
QY 118 TCHRDKNCIINKNVTRNRCQYCRLOKCFEVMGSKESVNRDRNKKKPKESYTLTP 177  
Db 116 TCHRDKNVINKVTRNRCQYCRLOKCFEVMGSKESVNRDRNKKKPKESQKCTSEYEMTA 175  
QY 178 EVGELIEKVRKAHOETFPALCQKGYTTNNSSQSVSLDLDLWDFSELSTKCIKIVF 237  
Db 176 ELDDLTEKIRKAHOETFPALCQKGYTTNNSSQSVSLDLDLWDFSELSTKCIKIVF 235  
QY 238 AKQLPGFTTLTIADQITLLKAAACLDILILRICTRTYTPQDTMTFSDGLTLNRTQMHNAGF 297  
Db 236 AKRLPGFTTLTIADQITLLKAAACLDILILRICTRTYTPQDTMTFSDGLTLNRTQMHNAGF 295  
QY 298 GPLTDLVAFANOLLPLEMDDAETGLLSAICLCIGDRQDLEOPDRVDMLEALVKY 357  
Db 296 GPLTDLVTFANOLLPLEMDDETETGLLSAICLCIGDRQDLEETPKVDKLOEPLLEALKIY 355  
QY 358 VRKRSPRPHFPMKMLKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDT 417  
Db 356 IRKRSPKPHFPMKMLKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDT 415  
QY 418 LSGOPGGGRDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
Db 416 LTPSSSGNIAEH-----SPSVSPSSVNSGVQSOP 445

## RESULT 13

retinoic acid receptor beta-4 - chicken  
C:Species: Gallus gallus (chicken)  
C>Date: 11-Dec-1992 #sequence\_revision 30-Jan-1993 #text\_change 20-Aug-1999  
C:Accession: A43786; S14291; A60129; S17072; S16243  
R:Smith, S.M.; Eichele, G.  
A:Title: Temporal and regional differences in the expression pattern of distinct retinoic acid receptor beta-4 isoforms in the chicken embryo.  
A:Reference number: A43786; MUID:91199956; PMID:1849811  
A:Accession: A43786  
A:Molecule type: mRNA  
A:Residues: 1-448 <SMI>  
A:Cross-references: GB:X59473; NID:963753; PIDN:CAA42077.1; PID:g63754  
A:Note: The authors translated the codon ACT for residue 96 as Ser, TCC for residue 167  
R:Noji, S.; Nohno, T.; Koyama, E.; Muto, K.; Ohya, K.; Aoki, Y.; Tamura, K.; Ohsugi, K.  
Nature 350, 83-86, 1991  
A:Title: Retinoic acid induces polarizing activity but is unlikely to be a morphogen in the chicken embryo.  
A:Reference number: S14291; MUID:91163640; PMID:1848357  
A:Accession: S14291  
A:Molecule type: mRNA  
A:Residues: 1-95, 'S', '97-448 <NOJ>  
A:Cross-references: EMBL:X57340; NID:g62976; PIDN:CAA40616.1; PID:g62977  
R:Rowe, A.; Richman, J.M.; Brickell, P.M.  
Development 111, 1007-1016, 1991  
A:Title: Retinoic acid treatment alters the distribution of retinoic acid receptor-beta isoforms in the chicken embryo.  
A:Reference number: A60129; MUID:91347912; PMID:1652423  
A:Accession: A60129  
A:Molecule type: mRNA  
A:Residues: 23-24, 26-95, 'S', '97-165, 'D', '167-448 <ROW>  
A:Cross-references: GB:S63196; NID:g234370; PIDN:AAB19628.1; PID:g234371  
R:Nohno, T.  
submitted to the EMBL Data Library, January 1991  
A:Reference number: S17072  
A:Accession: S17072  
A:Molecule type: mRNA  
A:Residues: 15-95, 'S', '97-434 <NOH>  
A:Cross-references: EMBL:X57339  
R:Nohno, T.; Muto, K.; Noji, S.; Saito, T.; Taniguchi, S.  
Biochim. Biophys. Acta 1089, 273-275, 1991  
A:Title: Isoforms of retinoic acid receptor beta expressed in the chicken embryo.

A:Reference number: S16243; MUID:91274365; PMID:1647216  
A:Accession: S16243  
A:Molecule type: mRNA  
A:Residues: 15-60 <NOH2>  
A:Cross-references: EMBL:X57339  
A:Superfamily: retinoic acid receptor alpha; erba transforming protein homology  
C:Keywords: alternative splicing; DNA binding; zinc finger  
F:79-330/Domain: erba transforming protein homology <ERBA>  
F:81-101/Region: zinc finger  
F:117-141/Region: zinc finger

Query Match 72.1%; Score 1757; DB 2; Length 448;  
Best Local Similarity 74.3%; Pred. No. 1.9e-121;  
Matches 342; Conservative 42; Mismatches 58; Indels 18; Gaps 5;

QY 1 MYESVEVGPTNPPLVVDVFNQNRACLLPEKGLPAFCPYSTPLRPLW--NGSNHSIE 57  
Db 1 MFDCMDVLSVSPGQ--ILDEVTASPSQMLQEKALKACLSGFTQAE--WQHRHTAQSI 55  
QY 58 TQSSSEIIVSPSPPLPRIVKPCFCQDKSSGYHYGVSAACGCGKGFRRSIQKNMY 117  
Db 56 TQSTSSSEIIVSPSPPLPRIVKPCFCQDKSSGYHYGVSAACGCGKGFRRSIQKNMY 115  
QY 118 TCHRDKNCIINKNVTRNRCQYCRLOKCFEVMGSKESVNRDRNKKKPKESYTLTP 177  
Db 116 TCHRDKNVINKVTRNRCQYCRLOKCFEVMGSKESVNRDRNKKKPKESQKCTSEYEMTA 175  
QY 178 EVGELIEKVRKAHOETFPALCQKGYTTNNSSQSVSLDLDLWDFSELSTKCIKIVF 237  
Db 176 ELDDLTEKIRKAHOETFPALCQKGYTTNNSSQSVSLDLDLWDFSELSTKCIKIVF 235  
QY 238 AKQLPGFTTLTIADQITLLKAAACLDILILRICTRTYTPQDTMTFSDGLTLNRTQMHNAGF 297  
Db 236 AKRLPGFTTLTIADQITLLKAAACLDILILRICTRTYTPQDTMTFSDGLTLNRTQMHNAGF 295  
QY 298 GPLTDLVAFANOLLPLEMDDAETGLLSAICLCIGDRQDLEOPDRVDMLEALVKY 357  
Db 296 GPLTDLVTFANOLLPLEMDDETETGLLSAICLCIGDRQDLEETPKVDKLOEPLLEALKIY 355  
QY 358 VRKRSPRPHFPMKMLKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDT 417  
Db 356 IRKRSPKPHFPMKMLKITDLRSISAKGAERVITLKEIPGSMPLIQEMLNSEGLDT 415  
QY 418 LSGOPGGGRDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
Db 416 LTPSSSGNIAEH-----SPSVSPSSVNSGVQSOP 445

## RESULT 14

retinoic acid receptor beta-1 - chicken  
C:Species: Gallus gallus (chicken)  
C>Date: 21-Nov-1993 #sequence\_revision 22-Apr-1995 #text\_change 15-Oct-1999  
C:Accession: S13512; S17073  
R:Padanilam, B.J.; McLeod, L.B.; Suzuki, H.; Solursh, M.  
Nucleic Acids Res. 19, 395, 1991  
A:Title: Nucleotide sequence of an isoform of chicken retinoic acid binding protein-1  
A:Reference number: S13512; MUID:91195065; PMID:1849630  
A:Accession: S13512  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-455 <PAD>  
A:Cross-references: EMBL:X56674; NID:g63785; PIDN:CAA39997.1; PID:g63786  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, November 1991  
R:Nohno, T.; Muto, K.; Noji, S.; Saito, T.; Taniguchi, S.  
Biochim. Biophys. Acta 1089, 273-275, 1991  
A:Title: Isoforms of retinoic acid receptor beta expressed in the chicken embryo.  
A:Reference number: S16243; MUID:91274365; PMID:1647216  
A:Accession: S17073  
A:Molecule type: mRNA  
A:Residues: 1-29, 'P', '31-80 <NOH>  
A:Cross-references: EMBL:X57339  
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology  
C:Keywords: alternative splicing; DNA binding; nucleus; transcription regulation; zli

F:86-337/Domain: erba transforming protein homology <ERBA>  
F:88-108/Region: zinc finger  
F:124-148/Region: zinc finger

Query Match 71.18; Score 1733; DB 2; Length 455;  
Best Local Similarity 77.18; Pred. No. 1.1e-119;  
Matches 330; Conservative 33; Mismatches 41; Indels 24; Gaps 2;

QY 30 PEKGLPAPGVTPLRTPLWNGSNHSETQSSSEIEIVPSPPLPRIVKPCFVCODK 89  
DB 49 PTSCSTPSP-----ATVETQSTSEELVPSPPLPRIVKPCFVCODK 94  
QY 90 SSGHYGVSAACEGCKGFFRRSIQKNMVTCHRDKNCIINKVTRNRCQYCRLOKCFEVGMS 149  
DB 95 SSGHYGVSAACEGCKGFFRRSIQKNMVTCHRDKNCIINKVTRNRCQYCRLOKCFEVGMS 154  
QY 150 KESVRNDRNKKKKEVPKESSEYTLTPVEGELIEKVRKAHOETFPALCQLGKTYTNNSS 209  
DB 155 KESVRNDRNKKKKEVPKESSEYTLTPVEGELIEKVRKAHOETFPALCQLGKTYTNNSS 214  
QY 210 EQRVSLDLDLWDFSELSTKCIITVEFAKOLPGFTTLTIADQITLLKAACDILILIRIC 269  
DB 215 DHRVRLDLGWDFSELAATKCIIVKFAKRLPGFTSLTIADQITLLKAACDILILIRIC 274  
QY 270 TRYPTQDTMTFSDGLTLNRTQMHNAGFGLTDLVFAFANQLLPLEMDDAETGLLSAICL 329  
DB 275 TRYPTQDTMTFSDGLTLNRTQMHNAGFGLTDLVFAFANQLLPLEMDDAETGLLSAICL 334  
QY 330 ICGDRQLEQPDVDMLOEPLLEALKYVRRKRPSPHMFPMKMLKITDLSISAKGAER 389  
DB 335 ICGDRQLEQPDVDMLOEPLLEALKYVRRKRPSPHMFPMKMLKITDLSISAKGAER 394  
QY 390 VITLKEIPGSMPLIOEMLENSEGLTSLQPGGGGDRGGGLPPPGSCSPSLSPSSNR 449  
DB 395 VITLKEIPGSMPLIOEMLENSEGLTSLQPGGGGDRGGGLPPPGSCSPSLSPSSNR 454  
QY 450 SSPATHSP 457  
DB 445 NSSVSQSP 452

RESULT 15  
I51257  
retinoic acid receptor gamma - zebra fish  
C:Species: Brachydanio rerio (zebra fish)  
C:Date: 04-Sep-1997 #sequence\_revision 04-Sep-1997 #text\_change 20-Aug-1999  
C:Accession: I51257  
R:Joore, J.; van der Lans, G.B.; Lanser, P.H.; Vervaaert, J.M.; Zivkovic, D.; Speksnijder  
Mech. Dev. 46, 137-150, 1994  
A:Title: Effects of retinoic acid on the expression of retinoic acid receptors during ze  
A:Reference number: I51256; MUID:95001557; PMID:7918098  
A:Accession: I51257  
A:Status: preliminary;  
A:Molecule type: mRNA  
A:Residues: 1-499 <JOO>  
A:Cross-references: GB:S74156; NID:g704371; PIDN:AAB32277.1; PID:g704372  
C:Genetics:  
A:Gene: zRARgamma  
C:Superfamily: retinoic acid receptor alpha; erba transforming protein homology  
F:75-327/Domain: erba transforming protein homology <ERBA>

Query Match 68.38; Score 1663.5; DB 2; Length 499;  
Best Local Similarity 71.38; Pred. No. 1.6e-114;  
Matches 328; Conservative 37; Mismatches 82; Indels 13; Gaps 6;

QY 1 MYESVEVGPTNPFLVDFYQNRACLLPEKGLPAPGYPSTPLRTPLWNG--SNHSIET 58  
DB 1 MFCMREALGMGRQ--LYDTNKG-ACMLRK-----ASPFYAGLDPPAWTGTASVRSVET 52  
QY 59 QSSSSEIEIVPSPPLPRIVKPCFVCODKSSGYHGVSAACEGCKGFFRRSIQKNMVT 118  
DB 53 QSTSSSEEMVPSPPPPPPRVKPCFVCODKSSGYHGVSAACEGCKGFFRRSIQKNMVT 112

QY 119 CHRDNKNCIINKVTRNRCQYCRLOKCFEVGMSKESVRNDRNKKKKEVPKPEC--SESYTLT 176  
DB 113 CHRDNKNCIINKVTRNRCQYCRLOKCFEVGMSKESVRNDRNKKKKEVPKPEC--SESYTLT 171  
QY 177 PEVGELIEKVRKAHOETFPALCQLGKTYTNNSSQORVSLDLDLWDFSELSTKCIITVE 236  
DB 172 GELELVNKKVSKAHOETFPALCQLGKTYTNNSSQORVSLDLDLWDFSELSTKCIITVE 231  
QY 237 FAKOLPGFTTLTIADQITLLKAACDILILIRICTRYTTPQDQMTFSDGLTLNRTQMHNAG 296  
DB 232 FAKRULPGFTTLTIADQITLLKAACDILILIRICTRYTTPQDQMTFSDGLTLNRTQMHNAG 291  
QY 297 FGPLTDLVFAFANQLLPLEMDDAETGLLSAICLICGDRQLEQPDVDMLOEPLLEALKV 356  
DB 292 FGPLTDLVFAFAGQLLPLEMDDTETGLLSAICLICGDRMDLEPERVDRLOEPLLEALKI 351  
QY 357 YVRRKRPSPHMFPMKMLKITDLSISAKGAERVITLKEIPGSMPLIOEMLENSEGLD 416  
DB 352 YARRRRPNKPHMFPMLMKITDLRGISTKGAEARITLKEIPGMPPLIREMLENPEAFE 411  
QY 417 TLSGPGGGGDRGGGLPPPGSCSPSLSPSSNRSPATHS 456  
DB 412 DQSEATEKKPEPEPPAPPALLTMKKEQDEDDSWATEN 451

Search completed: March 29, 2003, 09:43:58  
Job time : 37 secs

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 29, 2003, 06:37:58 ; Search time 24 Seconds  
(without alignments)  
560.261 million cell updates/sec

Title: US-09-691-220-2  
Perfect score: 2437  
Sequence: 1 MYESVEVGGPTNPFLLVDF.....SCSPSLSPSSNRSSPATHSP 457

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2127	87.3	462	2	US-08-592-383-2
2	2127	87.3	462	2	US-08-095-7288-4
3	2127	87.3	462	5	PCT-US92-02320A-4
4	2114	86.7	416	4	US-08-764-870-4
5	2114	86.7	416	4	US-08-980-115-4
6	2112	86.7	797	2	US-08-095-7288-2
7	2112	86.7	797	5	PCT-US92-02320A-2
8	2111	86.6	462	6	5171671-2
9	1816	74.5	403	2	US-08-592-383-4
10	1777	72.9	448	6	5223606-2
11	1662	68.2	454	4	US-08-764-870-5
12	1662	68.2	454	4	US-08-980-115-5
13	1662	68.2	454	6	5260432-2
14	1625	66.7	368	6	5223606-3
15	658	27.0	410	4	US-08-764-870-1
16	658	27.0	410	4	US-08-980-115-1
17	652	26.8	410	4	US-08-764-870-2
18	652	26.8	410	4	US-08-980-115-2
19	652	26.8	410	6	5438126-2
20	619.5	25.4	461	4	US-08-764-870-3
21	619.5	25.4	461	4	US-08-980-115-3
22	610	23.1	355	6	5223606-4
23	570	23.4	463	1	US-08-336-408B-6
24	570	23.4	463	5	PCT-US91-00399-6
25	562	23.1	460	1	US-08-342-411A-2
26	560	23.0	461	1	US-08-330-518-2
27	560	23.0	461	1	US-08-330-283-2

Sequence 2, Appli  
Sequence 2, Appli  
Sequence 2, Appli  
Sequence 1, Appli  
Sequence 8, Appli  
Sequence 8, Appli  
Sequence 2, Appli  
Sequence 3, Appli  
Sequence 3, Appli  
Sequence 4, Appli  
Sequence 7, Appli  
Sequence 7, Appli  
Sequence 2, Appli  
Sequence 4, Appli  
Sequence 2, Appli  
Sequence 6, Appli  
Sequence 6, Appli

28 560 23.0 451 2 US-08-646-248-2  
29 560 23.0 451 5 PCT-US95-13924-2  
30 560 23.0 461 1 PCT-US95-13931-2  
31 554 22.7 447 1 US-08-373-935-1  
32 552.5 22.7 440 1 US-08-333-358-8  
33 552.5 22.7 440 1 US-08-463-694-8  
34 552.5 22.7 440 1 US-08-694-501-8  
35 549 22.5 446 4 US-08-776-844-2  
36 546 22.4 446 2 US-08-372-652-3  
37 546 22.4 446 5 PCT-US95-16311-3  
38 545.5 22.4 443 1 US-08-342-411A-4  
39 542 22.2 525 4 US-08-764-870-7  
40 542 22.2 525 4 US-08-980-115-7  
41 531 21.8 533 1 US-07-952-800-2  
42 526.5 21.6 446 1 US-07-952-800-4  
43 518.5 21.3 462 1 US-08-336-408B-2  
44 518.5 21.3 462 4 US-08-764-870-6  
45 518.5 21.3 462 4 US-08-980-115-6

ALIGNMENTS

RESULT 1  
US-08-592-383-2  
; Sequence 2, Application US/08592383  
; Patent No. 5830760  
; GENERAL INFORMATION:  
; APPLICANT: Tsai, S. and S.J. Collins  
; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acid Re  
; NUMBER OF SEQUENCES: 10  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness  
; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue  
; CITY: Seattle  
; STATE: Washington  
; COUNTRY: USA  
; ZIP: 98101  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette-5.25 inch, 1.2mb storage  
; COMPUTER: IBM PC/386 Compatible  
; OPERATING SYSTEM: MS-DOS 4.01  
; SOFTWARE: Word for Windows 5.01-t  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/592,383  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/099,242  
; FILING DATE: July 28, 1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Broderick, Thomas F.  
; REGISTRATION NUMBER: 31,332  
; REFERENCE/DOCKET NUMBER: FPCR-1-7190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 462 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: polypeptide  
; DESCRIPTION: sequence encoded by SEQ. ID. NO. 1  
US-08-592-383-2

Query Match 87.3%; Score 2127; DB 2; Length 462;  
Best Local Similarity 91.6%; Pred. No. 1.3e-189;  
Matches 413; Conservative 5; Mismatches 13; Indels 20; Gaps 4;  
Qy 7 VGGTPNPFLLVDFYQNRACLLPEKGLPAGPPYSTPLRTPLNWNGSNHSIETQSSSSEI 66  
Db 32 LGLLSP-PGALTTLQHO-----LPVSG-YSTP-----SPATITQSSSSEI 71

QY 67 VSPSPPPPLPRIYKPCFCVQDKSSGYHYGVSACGCKGFFRRSIQKNMYTCHRDKNCI 126  
Db 72 VSPSPPPPLPRIYKPCFCVQDKSSGYHYGVSACGCKGFFRRSIQKNMYTCHRDKNCI 131  
QY 127 INKVTNRCOYCLQKCFEVMGSKESVRNDRNKKKEVPKPECSYTLTPVEGELIEKV 186  
Db 132 INKVTNRCOYCLQKCFEVMGSKESVRNDRNKKKEVPKPECSYTLTPVEGELIEKV 191  
QY 187 RKAHQETFPALCOLGKYTTNNSEORVSLDLDLWDFSELSKCIITKTVFAKOLPGFTT 246  
Db 192 RKAHQETFPALCOLGKYTTNNSEORVSLDLDLWDFSELSKCIITKTVFAKOLPGFTT 251  
QY 247 LTIADQITLLKAAACLDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFPLTDLVFA 306  
Db 252 LTIADQITLLKAAACLDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFPLTDLVFA 311  
QY 307 FANQLLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEALKYVYVRRRPSRP 366  
Db 312 FANQLLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEALKYVYVRRRPSRP 371  
QY 367 HMPKMLKITDLRSISAKGAERVITLKMIEPGSMPLIOEMLENSEGLDTLGQPGGGG 426  
Db 372 HMPKMLKITDLRSISAKGAERVITLKMIEPGSMPLIOEMLENSEGLDTLGQPGGGG 431  
QY 427 RDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
Db 432 RDGGGLAPPGGSCPSLSPSSNRSSPATHSP 462  
RESULT 2  
US-08-095-728B-4  
; Sequence 4, Application US/08095728B  
; Patent No. 5843642  
; GENERAL INFORMATION:  
; APPLICANT: DMITROVSKY, ETHAN  
; APPLICANT: WARRELL JR, RAYMOND P  
; APPLICANT: MILLER JR, WILSON H  
; APPLICANT: FRANKEL, STANLEY  
; TITLE OF INVENTION: METHODS FOR THE DETECTION AND  
; TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA (APL)  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOPER & DUNHAM LLP  
; STREET: 1185 AVENUE OF THE AMERICAS  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/095,728B  
; FILING DATE: 21-JUL-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/673,838  
; FILING DATE: 22-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WHITE, JOHN P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 38694-A  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 278-0400  
; TELEFAX: (212) 391-0525  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 462 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear

; MOLECULE TYPE: protein  
US-08-095-728B-4  
Query Match 87.3%; Score 2127; DB 2: Length 462;  
Best Local Similarity 91.6%; Pred. No. 1.3e-189;  
Matches 413; Conservative 5; Mismatches 13; Indels 20; Gaps 4;  
QY 7 VGGTTPNPFVLVDYFYNQNRACLLPEKGLPAPGYPYSTPLRTPLWNGSNHSIETQSSSSEI 66  
Db 32 LGGLSP-PGALTTLQHQ-----LPVSG-YSTP-----SPATIEQTSSSSEI 71  
QY 67 VPSPSPPPPLPRIYKPCFCVQDKSSGYHYGVSACGCKGFFRRSIQKNMYTCHRDKNCI 126  
Db 72 VPSPSPPPPLPRIYKPCFCVQDKSSGYHYGVSACGCKGFFRRSIQKNMYTCHRDKNCI 131  
QY 127 INKVTNRCOYCLQKCFEVMGSKESVRNDRNKKKEVPKPECSYTLTPVEGELIEKV 186  
Db 132 INKVTNRCOYCLQKCFEVMGSKESVRNDRNKKKEVPKPECSYTLTPVEGELIEKV 191  
QY 187 RKAHQETFPALCOLGKYTTNNSEORVSLDLDLWDFSELSKCIITKTVFAKOLPGFTT 246  
Db 192 RKAHQETFPALCOLGKYTTNNSEORVSLDLDLWDFSELSKCIITKTVFAKOLPGFTT 251  
QY 247 LTIADQITLLKAAACLDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFPLTDLVFA 306  
Db 252 LTIADQITLLKAAACLDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFPLTDLVFA 311  
QY 307 FANQLLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEALKYVYVRRRPSRP 366  
Db 312 FANQLLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEALKYVYVRRRPSRP 371  
QY 367 HMPKMLKITDLRSISAKGAERVITLKMIEPGSMPLIOEMLENSEGLDTLGQPGGGG 426  
Db 372 HMPKMLKITDLRSISAKGAERVITLKMIEPGSMPLIOEMLENSEGLDTLGQPGGGG 431  
QY 427 RDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
Db 432 RDGGGLAPPGGSCPSLSPSSNRSSPATHSP 462  
RESULT 3  
PCT-US92-02320A-4  
; Sequence 4, Application PC/TUS9202320A  
; GENERAL INFORMATION:  
; APPLICANT: Sloan-Kettering Institute, For Cancer Research  
; TITLE OF INVENTION: METHODS FOR DETECTION AND TREATMENT OF CANCER  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: COOPER & DUNHAM  
; STREET: 30 ROCKEFELLER PLAZA  
; CITY: NEW YORK  
; STATE: NEW YORK  
; COUNTRY: U.S.A.  
; ZIP: 10112  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.24  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US92/02320A  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 675,084  
; FILING DATE: 22-MAR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 673,838  
; FILING DATE: 22-MAR-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: WHITE, JOHN P  
; REGISTRATION NUMBER: 28,678  
; REFERENCE/DOCKET NUMBER: 38694-PCT





APPLICANT: Kushner, Peter J.  
APPLICANT: Apriletti, James W.  
APPLICANT: West, Brian L.  
APPLICANT: Shlau, Andrew K.  
TITLE OF INVENTION: NUCLEAR RECEPTOR LIGANDS AND LIGAND BINDING DOMAINS  
FILE REFERENCE: UCAL-246/0205  
CURRENT APPLICATION NUMBER: US/08/980,115  
CURRENT FILING DATE: 1997-11-26  
EARLIER APPLICATION NUMBER: 08/764,870  
EARLIER FILING DATE: 1996-12-13  
EARLIER APPLICATION NUMBER: 60/008,606  
EARLIER FILING DATE: 1995-12-14  
EARLIER APPLICATION NUMBER: 60/008,543  
EARLIER FILING DATE: 1995-12-13  
EARLIER APPLICATION NUMBER: 60/008,540  
EARLIER FILING DATE: 1995-12-13  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 4  
LENGTH: 416  
TYPE: PRT  
ORGANISM: Homo sapiens  
NAME/KEY: DOMAIN  
LOCATION: (131)...(373)  
OTHER INFORMATION: minimal ligand binding domain  
US-08-980-115-4

Query Match 86.7%; Score 2114; DB 4; Length 416;  
Best Local Similarity 97.1%; Pred. No. 1.8e-188;  
Matches 402; Conservative 2; Mismatches 2; Indels 8; Gaps 1;

QY 52 SNH-----SIETQSSSEIEIVSPSPPLPRIYKPCFVCDKSSGYHYGSACEG 103  
DB 3 SNHVASGAGEAAIETQSSSEIEIVSPSPPLPRIYKPCFVCDKSSGYHYGSACEG 62

QY 104 KGFRRSIQNMVYTCRDKNCIINKVTRNPOCYCRLOKCFEVMGSKESVRNDRNKKKE 163  
DB 63 KGFRRSIQNMVYTCRDKNCIINKVTRNPOCYCRLOKCFEVMGSKESVRNDRNKKKE 122

QY 164 VPKESESITLPEVGELEIKVRAHQETFPALCQLGKVTNNSSQORVSLDIDLWDF 223  
DB 123 VPKESESITLPEVGELEIKVRAHQETFPALCQLGKVTNNSSQORVSLDIDLWDF 182

QY 224 SELSTKCIITVFPAKOLPGFTLTADQITLLKAACDLILILICTRYPTQDTWTFSD 283  
DB 183 SELSTKCIITVFPAKOLPGFTLTADQITLLKAACDLILILICTRYPTQDTWTFSD 242

QY 284 GLTLNRTOMNAGFGPLTDLVFANOLLPLEMDDAETGLLSAICLCGRDQDLEQPDV 343  
DB 243 GLTLNRTOMNAGFGPLTDLVFANOLLPLEMDDAETGLLSAICLCGRDQDLEQPDV 302

QY 344 DMLQEPLEALKVYVRRRSPRPHMPKMLKITDLSISAKGAERVITLTKMEIPGSMPP 403  
DB 303 DMLQEPLEALKVYVRRRSPRPHMPKMLKITDLSISAKGAERVITLTKMEIPGSMPP 362

QY 404 LIQEMLENSGLDLSQCPGGGDDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
DB 363 LIQEMLENSGLDLSQCPGGGDDGGGLPPPGSCPSLSPSSNRSSPATHSP 416

RESULT 6  
US-08-095-7288-2  
Sequence 2, Application US/08095728B  
Patent No. 5843642  
GENERAL INFORMATION:  
APPLICANT: DMITROVSKY, ETHAN  
APPLICANT: WARRELL JR, RAYMOND P  
APPLICANT: MILLER JR, WILSON H  
APPLICANT: FRANKEL, STANLEY  
TITLE OF INVENTION: METHODS FOR THE DETECTION AND  
TREATMENT OF ACUTE PROMYELOCYTIC LEUKEMIA (APL)  
NUMBER OF SEQUENCES: 11

CORRESPONDENCE ADDRESS:  
ADDRESSEE: COOPER & DUNHAM LLP  
STREET: 1185 AVENUE OF THE AMERICAS  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/095,728B  
FILING DATE: 21-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/673,838  
FILING DATE: 22-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: WHITE, JOHN P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 38694-A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 278-0400  
TELEFAX: (212) 391-0525  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 797 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-095-7288-2

Query Match 86.7%; Score 2112; DB 2; Length 797;  
Best Local Similarity 99.3%; Pred. No. 6.9e-188;  
Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 55 SIETQSSSEIEIVSPSPPLPRIYKPCFVCDKSSGYHYGSACEGKGFRRSIQKN 114  
DB 395 AIETQSSSEIEIVSPSPPLPRIYKPCFVCDKSSGYHYGSACEGKGFRRSIQKN 454

QY 115 MVYTCRDKNCIINKVTRNPOCYCRLOKCFEVMGSKESVRNDRNKKKEVPKPESESYT 174  
DB 455 MVYTCRDKNCIINKVTRNPOCYCRLOKCFEVMGSKESVRNDRNKKKEVPKPESESYT 514

QY 175 LTPEVGELEIKVRAHQETFPALCQLGKVTNNSSQORVSLDIDLWDFSELSTKCIIT 234  
DB 515 LTPEVGELEIKVRAHQETFPALCQLGKVTNNSSQORVSLDIDLWDFSELSTKCIIT 574

QY 235 VEFKQLPGFTLTADQITLLKAACDLILILICTRYPTQDTWTFSDGLTLNRTOMHN 294  
DB 575 VEFKQLPGFTLTADQITLLKAACDLILILICTRYPTQDTWTFSDGLTLNRTOMHN 634

QY 295 AGFPLDVLFAFANOLLPLEMDDAETGLLSAICLCGRDQDLEQPDVDMLEQPLEAL 354  
DB 635 AGFPLDVLFAFANOLLPLEMDDAETGLLSAICLCGRDQDLEQPDVDMLEQPLEAL 694

QY 355 KVVYVRRRSPRPHMPKMLKITDLSISAKGAERVITLTKMEIPGSMPPLIQEMLENSG 414  
DB 695 KVVYVRRRSPRPHMPKMLKITDLSISAKGAERVITLTKMEIPGSMPPLIQEMLENSG 754

QY 415 LDTLSGQPGGGGDDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
DB 755 LDTLSGQPGGGGDDGGGLPPPGSCPSLSPSSNRSSPATHSP 797

RESULT 7  
PCT-US92-02320A-2  
Sequence 2, Application PC/TUS9202320A  
GENERAL INFORMATION:  
APPLICANT: Sloan-Kettering Institute, For Cancer Research  
TITLE OF INVENTION: METHODS FOR DETECTION AND TREATMENT OF CANCER

NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: COOPER & DUNHAM  
STREET: 30 ROCKEFELLER PLAZA  
CITY: NEW YORK  
STATE: NEW YORK  
COUNTRY: U.S.A.  
ZIP: 10112  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US92/02320A  
FILING DATE:  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 675,084  
FILING DATE: 22-MAR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 673,838  
FILING DATE: 22-MAR-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: WHITE, JOHN P  
REGISTRATION NUMBER: 28,678  
REFERENCE/DOCKET NUMBER: 38694-PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 977-9550  
TELEFAX: (212) 644-0525  
TELEX: (212) 422523 COOP UI  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 797 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US92-02320A-2

Query Match 86.7%; Score 2112; DB 5; Length 797;

Best Local Similarity 99.3%; Pred. No. 6.9e-188;

Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 55 SIETQSSSEIIVSPSPPLPRIYKPCFVCQDKSSGYHYGVSAACECKGFFRRSIQKN 114  
:|||||  
DB 395 AIETQSSSEIIVSPSPPLPRIYKPCFVCQDKSSGYHYGVSAACECKGFFRRSIQKN 454  
QY 115 MVTCHRDKNCIINKVTNRNCOYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPECSSEY 174  
:|||||  
DB 455 MVTCHRDKNCIINKVTNRNCOYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPECSSEY 514  
QY 175 LTPVEGLIEKVRKAHQTETPALCOLGKYTTNNSEQRVSLDIDLWDFSELSTKCIKT 234  
:|||||  
DB 515 LTPVEGLIEKVRKAHQTETPALCOLGKYTTNNSEQRVSLDIDLWDFSELSTKCIKT 574  
QY 235 VEFKQLPGFTTLTADQITLLKAAACDLILRICTRYTPEDQTMFTSDGLTLNRTOMHN 294  
:|||||  
DB 575 VEFKQLPGFTTLTADQITLLKAAACDLILRICTRYTPEDQTMFTSDGLTLNRTOMHN 634  
QY 295 AGFGPLDLVAFANQLLPLEMDDAETGLLSAICLICGDRDLEQPDVDMLOEPLLEAL 354  
:|||||  
DB 635 AGFGPLDLVAFANQLLPLEMDDAETGLLSAICLICGDRDLEQPDVDMLOEPLLEAL 694  
QY 355 KVVYKRRPSRPHMFKMLKITDLRSISAKGAERVITLKMIEPGSMPLPQIOMLENSEG 414  
:|||||  
DB 695 KVVYKRRPSRPHMFKMLKITDLRSISAKGAERVITLKMIEPGSMPLPQIOMLENSEG 754  
QY 415 LDTLSGQPGGGGRDGLPPPGSCPSLSNSSRSPATHSP 457  
:|||||  
DB 755 LDTLSGQPGGGGRDGLPPPGSCPSLSNSSRSPATHSP 797

RESULT 8

5171671-2  
; Patent No. 5171671  
; APPLICANT: EVANS, RONALD M.; ONG, ESTELITA S.; SEGUI,  
; PRUDIMAR S.; THOMPSON, CATHERINE C.; UEMSONO, KAZUHIKO  
; GUGUERE, VINCENT  
; TITLE OF INVENTION: RETINOIC ACID RECEPTOR COMPOSITION  
; NUMBER OF SEQUENCES: 2  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/546,256  
; FILING DATE: 06-AUG-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 276,536  
; FILING DATE: 30-NOV-1988  
; APPLICATION NUMBER: 128,331  
; FILING DATE: 02-DEC-1987  
; SEQ ID NO: 2:  
; LENGTH: 462  
5171671-2

Query Match 86.6%; Score 2111; DB 6; Length 462;

Best Local Similarity 91.1%; Pred. No. 3.9e-188;

Matches 411; Conservative 5; Mismatches 15; Indels 20; Gaps 4;

QY 7 VGGTNPPLVVDVFNQNRACLLPEKGLPAGPYSPPLRTPPLWNGSNHSTETQSSSEI 66  
:|||||  
DB 32 LGGLSP-FCALTTLQHQ-----LVSG-YSTP-----SPATITQSSSEI 71  
QY 67 VSPSPPLPRIYKPCFVCQDKSSGYHYGVSAACECKGFFRRSIQKNMYTCHRDKNCI 126  
:|||||  
DB 72 VSPSPPLPRIYKPCFVCQDKSSGYHYGVSAACECKGFFRRSIQKNMYTCHRDKNCI 131  
QY 127 INKVTNRNCOYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPECSSEYTLTPVEGLIEKV 186  
:|||||  
DB 132 INKVTNRNCOYCRLOKCFEYGVMSKESVRNDRNKKKEVPKPECSSEYTLTPVEGLIEKV 191  
QY 187 RKAHQTETPALCOLGKYTTNNSEQRVSLDIDLWDFSELSTKCIKTVEFAKQLPGFTT 246  
:|||||  
DB 192 RKAHQTETPALCOLGKYTTNNSEQRVSLDIDLWDFSELSTKCIKTVEFAKQLPGFTT 251  
QY 247 LTIADQITLLKAAACDLILRICTRYTPEDQTMFTSDGLTLNRTOMHNAGFGLTDLVFA 306  
:|||||  
DB 252 LTIADQITLLKAAACDLILRICTRYTPEDQTMFTSDGLTLNRTOMHNAGFGLTDLVFA 311  
QY 307 FANQLLPLEMDDAETGLLSAICLICGDRDLEQPDVDMLOEPLLEALKYVVRKRRSRP 366  
:|||||  
DB 312 FANQLLPLEMDDAETGLLSAICLICGDRDLEQPDVDMLOEPLLEALKYVVRKRRSRP 371  
QY 367 HMFPMKMLKITDLRSISAKGAERVITLKMIEPGSMPLPQIOMLENSEGDLTSGQPGGG 426  
:|||||  
DB 372 HMFPMKMLKITDLRSISAKGAERVITLKMIEPGSMPLPQIOMLENSEGDLTSGQPGGG 431  
QY 427 RDGGGLPPPGSCPSLSNSSRSPATHSP 457  
:|||||  
DB 432 RDGGGLPPPGSCPSLSNSSRSPATHSP 462

RESULT 9

US-08-592-383-4

; Sequence 4, Application US/08592383

; Patent No. 5830760

; GENERAL INFORMATION:

; APPLICANT: Tsai, S. and S.J. Collins

; TITLE OF INVENTION: "Hematopoietic Cell Lines Bearing Altered Retinoic Acid Res

; NUMBER OF SEQUENCES: 10

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness

; STREET: 2800 Pacific First Centre, 1420 Fifth Avenue

; CITY: Seattle

; STATE: Washington

; COUNTRY: USA

; ZIP: 98101

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette-5.25 inch, 1.2Mb storage

COMPUTER: IBM PC/386 Compatible  
OPERATING SYSTEM: MS-DOS 4.01  
SOFTWARE: Word for Windows 5.01-t  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/592,383  
FILING DATE:

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/099,242

FILING DATE: July 28, 1993  
ATTORNEY/AGENT INFORMATION:

NAME: Broderick, Thomas F.  
REGISTRATION NUMBER: 31,332

REFERENCE/DOCKET NUMBER: FPCR-1-7190  
TELEPHONE: 1-206-682-8100; 1-206-224-0709(direct)

INFORMATION FOR SEQ ID NO. 4:  
SEQUENCE CHARACTERISTICS:

LENGTH: 403 amino acids  
TYPE: amino acid

STRANDEDNESS: single  
TOPOLOGY: linear

MOLECULE TYPE: polypeptide  
DESCRIPTION: sequence encoded by SEQ. ID. NO. 3; RAR-alpha403

US-08-592-383-4

Query Match 74.5%; Score 1816; DB 2; Length 403;  
Best Local Similarity 90.6%; Pred. No. 9e-161; 4;  
Matches 355; Conservative 5; Mismatches 12; Indels 20; Gaps

QY 7 VGGPTNPFLVDFYQNRACLLPEKLPAGPYSTPLRTPLWNGSNHSTQSSSEEEI 66  
DB 32 LGLSLP-PCALATTLOHQ-----LPVSG-YSTP-----SPATITQSSSEEEI 71  
QY 67 VPSPSPPLPRIYKPCFCVQDKSSGYHYGVSAECGCKGFFRRSIQKNMYTCHRDKNCI 126  
DB 72 VPSPSPPLPRIYKPCFCVQDKSSGYHYGVSAECGCKGFFRRSIQKNMYTCHRDKNCI 131  
QY 127 INKVTNRNCOYRLQKCFEYGVSKSVNRDNKKKKEVPKPECSYTLTPVEGELIEKV 186  
DB 132 INKVTNRNCOYRLQKCFEYGVSKSVNRDNKKKKEVPKPECSYTLTPVEGELIEKV 191  
QY 187 RKAHQETFPALCOLGKYTTNNSEQRVSLDIDLWDFSELSTKCIITKVEFAKOLPGFTT 246  
DB 192 RKAHQETFPALCOLGKYTTNNSEQRVSLDIDLWDFSELSTKCIITKVEFAKOLPGFTT 251  
QY 247 LTIADQITLLKAACLDILILIRICTRYTPQDQTMFTSDGLTLNRTQMHNAGFGPLTLVFA 306  
DB 252 LTIADQITLLKAACLDILILIRICTRYTPQDQTMFTSDGLTLNRTQMHNAGFGPLTLVFA 311  
QY 307 FANQLLPLEMDDAETGLLSAICLICGDRDLEQPDQDQDQDQDQDQDQDQDQDQDQDQ 366  
DB 312 FANQLLPLEMDDAETGLLSAICLICGDRDLEQPDQDQDQDQDQDQDQDQDQDQDQD 371  
QY 367 HMFPMKMLKITDLRSISAKGAERVITLKMIEP 398  
DB 372 HMFPMKMLKITDLRSISAKGAERVITLKMIEP 403

RESULT 10

5223606-2

Patent No. 5223606

APPLICANT: BLAUDIN DE THE, HUGHES; MARCHIO, AGNES; TIOLLAS, PIERRE; DEJEAN, ANNE

TITLE OF INVENTION: STEROID/THYROID HORMONE RECEPTOR-RELATED

PROTEIN INAPPROPRIATELY EXPRESSED IN HUMAN HEPATOCELLULAR CARCINOMA

NUMBER OF SEQUENCES: 11

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/134,130

FILING DATE: 17-DEC-1987

PRIOR APPLICATION DATA:

SEQ ID NO.2:

LENGTH: 448

5223606-2

Query Match 72.9%; Score 1777; DB 6; Length 448;

Best Local Similarity 74.6%; Pred. No. 4.5e-157;

Matches 343; Conservative 42; Mismatches 57; Indels 18; Gaps 5;

QY 1 MYESVEVGGPTNPFLVDFYQNRACLLPEKLPAGPYSTPLRTPLWNGSN--HSTE 57  
DB 1 MFCMDVLVSFGQ--ILDFTAPSSCMLEKALKA---CFSGLTQTEWQHRHTAQSLIE 55  
QY 58 TQSSSEEEIVPSPSPPLPRIYKPCFCVQDKSSGYHYGVSAECGCKGFFRRSIQKNMY 117  
DB 56 TQSTSSSEELVPSPLPPRPVYKPCFCVQDKSSGYHYGVSAECGCKGFFRRSIQKNMY 115  
QY 118 TCHRDKNCIINKVTNRNCOYRLQKCFEYGVSKSVNRDNKKKKEVPKPECSYTLTP 177  
DB 116 TCHRDKNCIINKVTNRNCOYRLQKCFEYGVSKSVNRDNKKKKEVPKPECSYTLTP 175  
QY 178 EVGELIEKVRKAHQETFPALCOLGKYTTNNSEQRVSLDIDLWDFSELSTKCIITKVEF 237  
DB 176 ELDDLTEKIRKAHQETFPALCOLGKYTTNNSEQRVSLDIDLWDFSELSTKCIITKVEF 235  
QY 238 AKOLPGFTTITADQITLLKAACLDILILIRICTRYTPQDQTMFTSDGLTLNRTQMHNAG 297  
DB 236 AKRLPGFTTITADQITLLKAACLDILILIRICTRYTPQDQTMFTSDGLTLNRTQMHNAG 295  
QY 298 GPLTDLVFAFANQLLPLEMDDAETGLLSAICLICGDRDLEQPDQDQDQDQDQDQ 357  
DB 296 GPLTDLVFAFANQLLPLEMDDAETGLLSAICLICGDRDLEQPDQDQDQDQDQDQ 355  
QY 358 VRKRRSRPHMFKMLKITDLRSISAKGAERVITLKMIEFGSMPLLOEMNSEGLDT 417  
DB 356 IRRRRSPKPHMFKMLKITDLRSISAKGAERVITLKMIEFGSMPLLOEMNSEGLDT 415  
QY 418 LSGOPGGGGRDGGGLPPPGSCPSLSPSSNRSSPATHSP 457  
DB 416 LTPSSSGNTAEH-----SPSISPVSSVGVSGSP 445

RESULT 11

US-08-764-870-5

Sequence 5, Application US/08764870

Patent No. 6236946

GENERAL INFORMATION:

APPLICANT: Scanlan, Thomas S

APPLICANT: Baxter, John D

APPLICANT: Flettrick, Robert J

APPLICANT: Wagner, Richard L

APPLICANT: Kushner, Peter J

APPLICANT: Apriletti, James W

APPLICANT: West, Brian

TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand

TITLE OF INVENTION: Binding Domains

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Cooley Godward

STREET: Five Palo Alto Square, 3000 El Camino Real

CITY: Palo Alto

STATE: CA

COUNTRY: USA

ZIP: 94306

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/764,870

FILING DATE: 13-DEC-1996

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/008,540

FILING DATE: 13-DEC-1995

```

1  CURRENT APPLICATION NUMBER: US/08/980,115
2  CURRENT FILING DATE: 1997-11-26
3  EARLIER APPLICATION NUMBER: 08/764,870
4  EARLIER FILING DATE: 1996-12-13
5  EARLIER APPLICATION NUMBER: 60/008,606
6  EARLIER FILING DATE: 1995-12-14
7  EARLIER APPLICATION NUMBER: 60/008,543
8  EARLIER FILING DATE: 1995-12-13
9  EARLIER APPLICATION NUMBER: 60/008,540
10 EARLIER FILING DATE: 1995-12-13
11 NUMBER OF SEQ ID NOS: 17
12 SOFTWARE: Patent In Ver. 2.0
13 SEQ ID NO 5
14     LENGTH: 454
15     TYPE: PRT
16 ORGANISM: Homo sapiens
17 FEATURE:
18 NAME/KEY: DOMAIN
19 LOCATION: (179)..(421)
20 OTHER INFORMATION: minimal ligand binding
21 US-08-980-115-5

```

Query Match	68.2%;	Score 1662;	DB 4;	Length 454;
Best Local Similarity	72.8%;	Pred. No. 2.3e-146;		
Matches 310.	Conservative	39.	Mismatches 48;	Indels 32;
				Gaps 4;

[illegible]

RESULT 13  
5260432-2  
; Patent No. 5260432  
; APPLICANT: TAKAKU, FUMIMARO; ISHIKAWA, TAKASHI; IMAWARI, MICHIO;  
; EVANS, RONALD M.; UMESONO, KAZUHIKO  
; TITLE OF INVENTION: HUMAN GAMMA RETINOIC ACID RECEPTOR DNA  
; ;  
; NUMBER OF SEQUENCES: 4  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/370,407  
; FILING DATE: 22-JUN-1989  
; SEQ ID NO:2:  
; LENGTH: 454  
5260432-2

Query Match 68.2%; Score 1662; DB 6; Length 454;  
Best Local Similarity 72.8%; Pred. No. 2.3e-146;  
Matches 319; Conservative 39; Mismatches 48; Indels 32; Gaps 4;

QY 30 PEKGLPAG-PYSTP-----LRTPLWNG-----SNHSETQSSSEIYV 68  
DB 16 PGSGYPCAGFPFAPGALRGSPFPEMLSPFRGLGQDLPKEMASLSVETQSTSEEMP 75  
QY 69 SPSPPPPLRIYKPCFCODKSSGYHYGVSAECGCKGFFRRSIQKNMYVYCHRDKNKIIN 128  
DB 76 SPSPPPPPRVYKPCFCVNDKSSGYHYGVSSCEGCKGFFRRSIQKNMYVYCHRDKNKIIN 135  
QY 129 KVTNRNCOYCLQKCEFGVSKESVRNDRNKKKKEVPKPCSESYTLTPEVGELIEKVRK 188  
DB 136 KVTNRNCOYCLQKCEFGVSKESVRNDRNKKKKEVPKPCSESYTLTPEVGELIEKVRK 195  
QY 189 AHQETFPALCOLGKYTTNNSEORVSLDIDLWDKFSLSKCIKTVFAKQLPGFTTLT 248  
DB 196 AHQETFPALCOLGKYTTNNSEORVSLDIDLWDKFSLSKCIKTVFAKQLPGFTTLT 255  
QY 249 IADQITLLKAACLDILIRICTRYTPQDTMTFSDGLTLNRTQHNAGFGLTDLVFAFA 308  
DB 256 IADQITLLKAACLDILIRICTRYTPQDTMTFSDGLTLNRTQHNAGFGLTDLVFAFA 315  
QY 309 NQLPLEMDDAETGLLSAICLICGDRDLPQDRVDMLOEPLEALKVYVYKRRPSRPHM 368  
DB 316 GQLPLEMDDTETGLLSAICLICGDRDLPQDRVDMLOEPLEALKVYVYKRRPSRPHM 375  
QY 369 FPKMLKITDLRISAKGAERVTTLKMEIPGSMPLLIQEMLENSEGLDITLSGQPG----- 423  
DB 376 FPKMLKITDLRISAKGAERVTTLKMEIPGSMPLLIQEMLENSEGLDITLSGQPG----- 435  
QY 424 -----GGGRDGGGLPPP 435  
DB 436 SSEDEVPGGGKGLKSP 453

RESULT 14  
5223606-3  
; PATENT NO. 5223606  
; APPLICANT: BLAUDIN DE THE, HUGHES, MARCHIO, AGNES, TIOLLAIS,  
; PIERRE, DEJEAN, ANNE  
; TITLE OF INVENTION: STEROID/THYROID HORMONE RECEPTOR-RELATED  
; PROTEIN INAPPROPRIATELY EXPRESSED IN HUMAN HEPATOCELLULAR CARCINOMA  
; NUMBER OF SEQUENCES: 11  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/134,130  
; FILING DATE: 17-DEC-1987  
; PRIOR APPLICATION DATA:  
; SEQ ID NO:3:  
; LENGTH: 368  
5223606-3

Query Match 66.7%; Score 1625; DB 6; Length 368;  
Best Local Similarity 81.6%; Pred. No. 4.6e-143;  
Matches 306; Conservative 27; Mismatches 32; Indels 10; Gaps 1;

QY 83 CFVQDKSSGYHYGVSAECGCKGFFRRSIQKNMYVYCHRDKNKIINKVTNRNCOYCLQK 142  
DB 1 CFVQDKSSGYHYGVSAECGCKGFFRRSIQKNMYVYCHRDKNKIINKVTNRNCOYCLQK 60  
QY 143 CFVGNKESVRNDRNKKKKEVPKPCSESYTLTPEVGELIEKVRKAHOETFPALCOLGK 202  
DB 61 CFVGNKESVRNDRNKKKKEVPKPCSESYTLTPEVGELIEKVRKAHOETFPALCOLGK 120  
QY 203 YTTNNSSEORVSLDIDLWDKFSLSKCIKTVFAKQLPGFTTLTDLVFAFANQLLPLEMDDAETG 262  
DB 121 YTTNNSSEORVSLDIDLWDKFSLSKCIKTVFAKQLPGFTTLTDLVFAFANQLLPLEMDDAETG 180  
QY 263 ILILIRICTRYTPQDTMTFSDGLTLNRTQHNAGFGLTDLVFAFANQLLPLEMDDAETG 322  
DB 181 ILILIRICTRYTPQDTMTFSDGLTLNRTQHNAGFGLTDLVFAFANQLLPLEMDDAETG 240

QY 323 LLSAICLICGDRDLPQDRVDMLOEPLEALKVYVYKRRPSRPHMFKMLKITDLRSI 382  
DB 241 LLSAICLICGDRDLPQDRVDMLOEPLEALKVYVYKRRPSRPHMFKMLKITDLRSI 300  
QY 383 SAKGAERVTTLKMEIPGSMPLLIQEMLENSEGLDITLSGQPGGGGRDGLPPPPGSCSPS 442  
DB 301 SAKGAERVTTLKMEIPGSMPLLIQEMLENSEGLDITLSGQPGGGGRDGLPPPPGSCSPS 350  
QY 443 LSPSSNRSSPATHSP 457  
DB 351 ISPSSVNSGVQSOP 365

RESULT 15  
US-08-764-870-1  
; Sequence 1, Application US/08764870  
; Patent No. 6236946  
; GENERAL INFORMATION:  
; APPLICANT: Scanlan, Thomas S  
; APPLICANT: Baxter, John D  
; APPLICANT: Fletcher, Robert J  
; APPLICANT: Wagner, Richard L  
; APPLICANT: Kushner, Peter J  
; APPLICANT: Apriletti, James W  
; APPLICANT: West, Brian  
; TITLE OF INVENTION: Nuclear Receptor Ligands and Ligand  
; TITLE OF INVENTION: Binding Domains  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Cooley Godward  
; STREET: Five Palo Alto Square, 3000 El Camino Real  
; CITY: Palo Alto  
; STATE: CA  
; COUNTRY: USA  
; ZIP: 94306  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/764,870  
; FILING DATE: 13-DEC-1996  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/008,540  
; FILING DATE: 13-DEC-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/008,543  
; FILING DATE: 13-DEC-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 60/008,606  
; FILING DATE: 14-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Nakamura, Jackie N  
; REGISTRATION NUMBER: 35,966  
; REFERENCE/DOCKET NUMBER: UCAL-246/01US  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (650)843-5000  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 410 amino acids  
; TYPE: amino acid  
; STRANDEDNESS:  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-764-870-1

Query Match 27.0%; Score 658; DB 4; Length 410;  
Best Local Similarity 38.1%; Pred. No. 5e-53;  
Matches 139; Conservative 64; Mismatches 120; Indels 42; Gaps 7;

QY 83 CFVQDKSSGYHYGVSAECGCKGFFRRSIQKNM--VYCHRDKNKIINKVTNRNCOYCLQK 140

```

Db 53 CVVCGDKATGYHYRCITCGSCGFFRTIQKNLHPYSCKYDSCCVIDKITRQCQLCRF 112
QY 141 QKCFEYVGHKSVRND-----RNKKKEY-----PKPECSSEYTLTPEV 179
Db 113 KCIAGVGMADLVLDOSKRVAKRKLTQNRERRRKEEMIRSLQQRPEP-----TPEE 164
QY 180 GELIEKVRKAHOETPA---LCQLGKYTTNNSSEQRV-----SLDIDLWDFSELSTK 229
Db 165 WDLIHVATEAHRSTNAQGSWHKORRFLPDDIGQSPIVSMPDGDVKVDLEAFSEFTKIITP 224
QY 230 CIIKTVEFAKQLPGFTTLTIADQITLLKAACLDILITRICTRYTPQDWTMTFSDGLTLNR 289
Db 225 AITRVVDFAKLPMESELPCEDQIILLKCCMEIMSLRAAVRYDPESDITLTSGEWTVKR 284
QY 290 TOMHNAGFGPLDVLVFAFANQLLPLEMDDAETGLLSAICLICGDRDQLEQPDVDMQLEP 349
Db 285 KOLKNGGLGVVSDAIFELGKLSAFNLDDTEVALLQAVLLMSTDRSGLLCVDKIEKSQEA 344
QY 350 LLEALKVYVRKRPSRPHMFKMLAKITDLRSISAKGAERVITLKMIEPGSM-PPLIQEM 408
Db 345 YLLAFEHYVNRKHNIPHFWPKLLMKVTDLRMIGACHASRFLHMKVECTELFPPLFLEV 404
QY 409 LENSE 413
Db 405 FEDQE 409
```

Search completed: March 29, 2003, 09:44:44  
Job time : 26 secs

## SUMMARIES

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	2437	100.0	457	23	AA021489		Human nuclear horm
2	2127	87.3	452	13	AA027534		RAR-alpha protein.
3	2127	87.3	462	20	AA081964		Human RAR-alpha pr
4	2124	87.2	462	16	AA068023		RAR-alpha. Homo s
5	2119	87.0	462	10	AA090395		Retinoic acid rece
6	2114	86.7	416	20	AA021636		Ligand binding dom
7	2114	86.7	547	22	AB018958		Novel human diagno
8	2112	86.7	593	22	AB020948		Novel human diagno
9	2112	86.7	668	22	AB018862		Novel human diagno
10	2112	86.7	797	13	AA027533		mv1/RAR-alpha fusi

## ALIGNMENTS

PT as models or targets

New human nuclear hormone receptor proteins and nucleic acids, useful as models or targets for developing human therapeutic targets, and in



PT Identifying therapeutic proteins and modulators of nuclear hormone  
PT receptor expression  
XX  
PS Claim 1; Fig 2; 73pp; English.  
PS  
CC The invention relates to an isolated peptide of a novel human nuclear  
CC hormone receptor with a fully defined sequence of 457 amino acids given  
CC in the specification. The novel human nuclear hormone receptor peptides  
CC and nucleic acids encoding them can be used as models for the development  
CC of human therapeutic targets, aid in the identification of therapeutic  
CC proteins, and serve as targets for the development of human therapeutic  
CC agents that modulate nuclear hormone receptor activity in cells and  
CC tissues that express the nuclear hormone receptor. The nucleic acids may  
CC be used as a query sequence to perform searches against sequence  
CC databases to identify family members or related sequences, as probes or  
CC primers, to construct recombinant vectors, to identify compounds that  
CC modulate nuclear hormone receptor nucleic acid expression, in gene  
CC therapy, and as antisense constructs to control nuclear hormone receptor  
CC gene expression in cells, tissues or organisms. The polypeptides can be  
CC used to raise antibodies or to elicit an immune response, as a reagent in  
CC assays designed to determine protein levels in biological fluids, as  
CC markers for tissues in which a corresponding protein is expressed, to  
CC identify a binding partner/ligand to develop a system for the  
CC identification of inhibitors of the binding reaction, in drug screening  
CC assays, and to identify compounds that modulate protein activity. This  
CC sequence represents the human nuclear hormone receptor protein of the  
CC invention.  
XX  
SQ Sequence 457 AA;

Query Match 100.0%; Score 2437; DB 23; Length 457;  
Best Local Similarity 100.0%; Pred. No. 7.7e-201;  
Matches 457; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MYEVEVGPTNPFLVDFYQNRACILPEKGLPAGPYSTPLRTPLNWSNHSIETQS 60  
Db 1 MYEVEVGPTNPFLVDFYQNRACILPEKGLPAGPYSTPLRTPLNWSNHSIETQS 60  
QY 61 SSSEIIVSPSPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRSIQKNMYTCH 120  
Db 61 SSSEIIVSPSPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRSIQKNMYTCH 120  
QY 121 RDKNCILINKVTRNPOCYCRLQKCFEVMGSKESVRNDRNKKKEVPKPECSYTLTPEVG 180  
Db 121 RDKNCILINKVTRNPOCYCRLQKCFEVMGSKESVRNDRNKKKEVPKPECSYTLTPEVG 180  
QY 181 ELIEKVRKAHQETFPALCOLGKYVTNNSSQORVSLDLDWDFSELSTKCIKTVEFAK 240  
Db 181 ELIEKVRKAHQETFPALCOLGKYVTNNSSQORVSLDLDWDFSELSTKCIKTVEFAK 240  
QY 241 LPGAFTLTADQITLLKAACTDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGL 300  
Db 241 LPGAFTLTADQITLLKAACTDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGL 300  
QY 301 TDLYFAFANQLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEALKYVVRK 360  
Db 301 TDLYFAFANQLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEALKYVVRK 360  
QY 361 RPSRPHFPKMLKITDURSTISAKGAERVITLKMIEPGSMPLIQEMLENSEGLDTLSG 420  
Db 361 RPSRPHFPKMLKITDURSTISAKGAERVITLKMIEPGSMPLIQEMLENSEGLDTLSG 420  
QY 421 QPGGGRGDGGGLPPPGGSCPSLSPSSNRSSPATHSP 457  
Db 421 QPGGGRGDGGGLPPPGGSCPSLSPSSNRSSPATHSP 457

RESULT 2  
AAR27534  
ID AAR27534 standard; Protein; 462 AA.  
XX  
AC AAR27534;  
XX

DT 09-MAR-1993 (first entry)  
XX  
DE RAR-alpha protein.  
XX  
KW Retinoic acid receptor; RAR-alpha; myl; acute promyelocytic leukemia;  
KW APL; translocation; chromosome 17; PCR; primer;  
KW [t(15;17)(q21;q11-22)]; breakpoint; polymerase chain reaction.  
XX  
OS Synthetic.  
XX  
PN W09216660-A.  
XX  
PD 01-OCT-1992.  
XX  
PF 23-MAR-1992; 92WO-US02320.  
XX  
PR 22-MAR-1991; 91US-0673838.  
PR 22-MAR-1991; 91US-0675084.  
XX  
(SLOK ) SLOAN KETTERING INST CANCER.  
XX  
PI Dmitrovsky E, Evans RM, Frankel S, Kazizuka A, Miller WH;  
PI Warrell RP;  
XX  
DR WPI; 1992-349240/42.  
DR N-PSDB; AAQ29338.  
XX  
PT Marker for acute promyelocytic leukaemia and other neoplasias -  
PT comprising nucleic acid and encoded abnormal retinoic acid  
PT receptor-alpha receptor  
XX  
PS Disclosure; Page 43-46; 84pp; English.  
XX  
CC The sequence given represents the amino acid sequence of the retinoic  
CC acid receptor (RAR)-alpha protein. The gene encoding this protein is  
CC disrupted in a translocation of a portion of the long arm of chromosome  
CC 17 onto the long arm of chromosome 15 [t(15;17)(q21;q11-22)]. This  
CC causes a fusion RAR-alpha and myl which is characteristic of acute  
CC promyelocytic leukemia (APL). The breakpoint region has been cloned  
CC and it has been shown that DNA rearrangements are clustered in the  
CC region of the first intron of RAR-alpha. This sequence was isolated  
CC by polymerase chain reaction (PCR). The primers used for amplification  
CC of this sequence can also be used to amplify the translocated region.  
XX  
SQ Sequence 462 AA;

Query Match 87.3%; Score 2127; DB 13; Length 462;  
Best Local Similarity 91.6%; Pred. No. 3.5e-174;  
Matches 413; Conservative 5; Mismatches 13; Indels 20; Gaps 4;  
QY 7 VGGPTNPFLVDFYQNRACILPEKGLPAGPYSTPLRTPLNWSNHSIETQSSEI 66  
Db 32 LGLSP-PGALTTLQHQ-----LPVSG-YSTP-----SPATITQSSEI 71  
QY 67 VPSPSPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRSIQKNMYTCHRDNCI 126  
Db 72 VPSPSPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRSIQKNMYTCHRDNCI 131  
QY 127 INKVTNPOCYCRLQKCFEVMGSKESVRNDRNKKKEVPKPECSYTLTPEVGELEK 186  
Db 132 INKVTNPOCYCRLQKCFEVMGSKESVRNDRNKKKEVPKPECSYTLTPEVGELEK 191  
QY 187 RKAHQETFPALCOLGKYVTNNSSQORVSLDLDWDFSELSTKCIKTVEFAKLPGETT 246  
Db 192 RKAHQETFPALCOLGKYVTNNSSQORVSLDLDWDFSELSTKCIKTVEFAKLPGETT 251  
QY 247 LTIADQITLLKAACTDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGLTDLVFA 306  
Db 252 LTIADQITLLKAACTDILILRICTRYTPEQDTMTFSDGLTLNRTQMHNAGFGLTDLVFA 311  
QY 307 FANQLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEALKYVVRKRSRP 366  
Db 312 FANQLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEALKYVVRKRSRP 371

QY 367 HMFPMKMKITDLRSISAKAERVITLKMIEIPGSMPLIQEMLENSEGLDITLSGQPGGGG 426  
Db 372 HMFPMKMKITDLRSISAKAERVITLKMIEIPGSMPLIQEMLENSEGLDITLSGQPGGGG 431  
QY 427 RDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457  
Db 432 RDGGGLPPPGSCSPSLSPSSNRSSPATHSP 462  
RESULT 3  
AAW81964  
ID AAW81964 standard; Protein; 462 AA.  
AC AAW81964;  
XX  
DT 05-FEB-1999 (first entry)  
DE Human RAR-alpha protein.  
KW Fusion protein; myl; retinoic acid receptor-alpha; RAR; human;  
KW acute promyelocytic leukaemia; APL; t(15;17); translocation;  
KW treatment; all-trans retinoic acid.  
XX  
OS Homo sapiens.  
XX  
XX US5843642-A.  
XX  
PD 01-DEC-1998.  
XX  
XX 21-JUL-1993; 93US-0095728.  
XX  
XX 21-JUL-1993; 93US-0095728.  
PR 22-MAR-1991; 91US-0673838.  
PR 22-MAR-1991; 91US-0675084.  
XX  
XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
XX  
XX Dmitrovsky E, Frankel S, Miller WH, Warrell RP;  
XX  
XX WPI; 1999-044563/04.  
DR N-PSDB; AAV64991.  
XX  
XX Diagnosis of acute promyelocytic leukaemia - by detecting nucleic  
PT acid encoding abnormal retinoic acid receptor-alpha  
XX  
XX Disclosure; Column 35-38; 38pp; English.  
XX  
XX This sequence represents the human retinoic acid receptor alpha,  
CC RAR-alpha which is used in a method for identifying a subject with acute  
CC promyelocytic leukaemia (APL) resulting from a t(15;17) translocation who  
CC will respond to treatment with all-trans retinoic acid. The protein can  
CC also be used to identify a subject with indications of APL who will not  
CC respond to treatment with all-trans retinoic acid.  
XX  
XX Sequence 462 AA;  
SQ  
Query Match 87.3%; Score 2127; DB 20; Length 462;  
Best Local Similarity 91.6%; Pred. No. 3.5e-174;  
Matches 413; Conservative 5; Mismatches 13; Indels 20; Gaps 4;  
QY 7 VGGTPPFLVVDYFNQNRACLLPEKGLPAPGPYSTPLRTPLWNGSNHSIETQSSSEI 66  
Db 32 LGLGSP-PGALTTLQH-----LPVSG-YSTP-----SPATIETQSSSEI 71  
QY 67 VPSPSPPLPRIKPCFVCGDKSSGYHGVSAACEGCKGFFRRSIQKNMVTCHRDKNCI 126  
Db 72 VPSPSPPLPRIKPCFVCGDKSSGYHGVSAACEGCKGFFRRSIQKNMVTCHRDKNCI 131  
QY 127 INKVTNRNCOYCRLOKCFEYGMKSRESYRNDNRNKKKKEVPKPECSSEYTLTPEVGELIEKV 186  
Db 132 INKVTNRNCOYCRLOKCFEYGMKSRESYRNDNRNKKKKEVPKPECSSEYTLTPEVGELIEKV 191

QY 187 RKAHQTFPALCOLGKYTTNNSSEORVSLDLDLWDFSELSKCIKTVFAKOLPGFTT 246  
Db 192 RKAHQTFPALCOLGKYTTNNSSEORVSLDLDLWDFSELSKCIKTVFAKOLPGFTT 251  
QY 247 LTIADQITLLKAACLDILILIRICTRYTPEQDTMTFSDDLTLNRTQMHNAGFGPLTDLVFA 306  
Db 252 LTIADQITLLKAACLDILILIRICTRYTPEQDTMTFSDDLTLNRTQMHNAGFGPLTDLVFA 311  
QY 307 FANQLLPLEMDDAETGLLSAICLICGDRDLEQPDVDMLOEPLLEALKVYVRRKRSRP 366  
Db 312 FANQLLPLEMDDAETGLLSAICLICGDRDLEQPDVDMLOEPLLEALKVYVRRKRSRP 371  
QY 367 HMFPMKMKITDLRSISAKAERVITLKMIEIPGSMPLIQEMLENSEGLDITLSGQPGGGG 426  
Db 372 HMFPMKMKITDLRSISAKAERVITLKMIEIPGSMPLIQEMLENSEGLDITLSGQPGGGG 431  
QY 427 RDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457  
Db 432 RDGGGLPPPGSCSPSLSPSSNRSSPATHSP 462  
RESULT 4  
AAR68023  
ID AAR68023 standard; Protein; 462 AA.  
AC AAR68023;  
XX  
DT 01-SEP-1995 (first entry)  
DE RAR-alpha.  
XX  
KW RAR-alpha; retinoic acid receptor alpha; hematopoietic; stem cell;  
KW differentiation.  
OS Homo sapiens.  
XX  
XX WO9504143-A.  
XX  
PD 09-FEB-1995.  
XX  
XX 28-JUL-1994; 94WO-US08450.  
XX  
XX 28-JUL-1993; 93US-0099242.  
PR (HUTC-) HUTCHINSON CANCER RES CENT FRED.  
XX  
XX Collins SJ, Tsai S;  
XX  
XX WPI; 1995-082227/11.  
DR N-PSDB; AAO81476.  
XX  
XX New haematopoietic stem cell lines with specific differentiation  
PT properties - made by transfecting stem cells with nucleic acid  
PT encoding dominant negative suppressor of the retinoic acid  
PT receptor alpha, useful e.g. for haematopoietic reconstitution  
XX  
XX Disclosure; Page 56-58; 100pp; English.  
XX  
XX An RAR-alpha cDNA (given in AAO81477) was prepared that contained a  
CC truncation of sequences coding for the C-terminal 59 amino acids  
CC and part the 3' UTR of the wild-type human sequence (AAO81476).  
CC This truncated cDNA, designated RAR-alpha-403, encodes a protein  
CC (AAR68024) containing the N-terminus, DNA-binding domain and part of  
CC the hormone-binding domain of RAR-alpha (AAR68023).  
XX  
XX Sequence 462 AA;  
SQ  
Query Match 87.2%; Score 2124; DB 16; Length 462;  
Best Local Similarity 91.4%; Pred. No. 6.4e-174;  
Matches 412; Conservative 6; Mismatches 13; Indels 20; Gaps 4;  
QY 7 VGGTPPFLVVDYFNQNRACLLPEKGLPAPGPYSTPLRTPLWNGSNHSIETQSSSEI 66  
Db 32 LGLGSP-PGALTTLQH-----LPVSG-YSTP-----SPATIETQSSSEI 71

Db 32 LGGLSP-PGALTTLQHQ-----LPVSG-YSTP-----SPATITQSSSEI 71

QY 67 VPSPPSPPLPRYKPCFVCQDKSSGYHYGSACEGCKGFFRRSIQKNMVTCHRDKNCI 126

Db 72 VPSPPSPPLPRYKPCFVCQDKSSGYHYGSACEGCKGFFRRSIQKNMVTCHRDKNCI 131

QY 127 INKVTNRCQYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPEVGELIEKV 186

Db 132 INKVTNRCQYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPEVGELIEKV 191

QY 187 RKAHQETFPALCOLGKYTTNNSSEQRVSLDIDLWDFSELSTKCIITVEFAKOLPGFTT 246

Db 192 RKAHQETFPALCOLGKYTTNNSSEQRVSLDIDLWDFSELSTKCIITVEFAKOLPGFTT 251

QY 247 LTIAQDITLLKAAACDILILIRICTRYTPQDQDLEQDPRVDMLOEPLEALKYVYKRRPSRP 306

Db 252 LTIAQDITLLKAAACDILILIRICTRYTPQDQDLEQDPRVDMLOEPLEALKYVYKRRPSRP 311

QY 307 FANQLLPLEMDDAETGLLSAICLICGDRDLEQDPRVDMLOEPLEALKYVYKRRPSRP 366

Db 312 FANQLLPLEMDDAETGLLSAICLICGDRDLEQDPRVDMLOEPLEALKYVYKRRPSRP 371

QY 367 HMFPMKMKITDLRSISAKGAERVITLKMIEPGSMPLIQEMLENSEGLDTLSGQPGGGG 426

Db 372 HMFPMKMKITDLRSISAKGAERVITLKMIEPGSMPLIQEMLENSEGLDTLSGQPGGGG 431

QY 427 RDGGLPPLPPGSCSPSLPSSNRSSPATHSP 457

Db 432 RDGGLAPPPGSCSPSLPSSNRSSPATHSP 462

RESULT 5

AAP90395

ID AAP90395 standard; protein; 462 AA.

XX AC AAP90395;

XX DT 01-NOV-1989 (first entry)

XX DE Retinoic acid receptor.

XX KW Clone phrAR1; retinoic acid receptor;

XX KW ligand complexes; human.

XX OS Homo sapiens (Human).

XX PN W08905355-A.

XX PD 15-JUN-1989.

XX PF 01-DEC-1988; 88WO-US04284.

XX PR 02-DEC-1987; 87US-0276536.

XX PA (SALK ) SALK INST FOR BIOLOGICAL STUD.

XX PI Evans RM, Giguere V, Ong ES, Segui PS, Umesono K, Thompson CC;

XX DR WPI; 1989-192701/26.

XX DR N-PSDB; AAN90124.

XX PT DNA encoding retinoic acid receptor proteins

XX PT - used to produce proteins for studying complexes with

XX PT ligands and in diagnostic assays.

XX PS Disclosure; fig 1b; 74pp; English.

XX CC Primary sequence of a protein that has ligand binding and

XX CC transcription activating properties of retinoic acid receptor

XX CC (RAR) protein (see corresp. AAN90124). Used to make chimeric

XX CC receptors, to produce receptor, to study binding complexes,

XX CC and to screen cpds. for RAR-agonists and antagonists.

SQ Sequence 462 AA;

Query Match 87.0%; Score 2119; DB 10; Length 462;

Best Local Similarity 91.4%; Pred. No. 1.7e-173;

Matches 412; Conservative 5; Mismatches 14; Indels 20; Gaps 4;

QY 7 VGGPTNPFLVDFYQNRACLLPEKGLPAGPYSTRLPTPLWNGSNHSIETQSSSSEI 66

Db 32 LGGLSP-PGALTTLQHQ-----LPVSG-YSTP-----SPATITQSSSEI 71

QY 67 VPSPPSPPLPRYKPCFVCQDKSSGYHYGSACEGCKGFFRRSIQKNMVTCHRDKNCI 126

Db 72 VPSPPSPPLPRYKPCFVCQDKSSGYHYGSACEGCKGFFRRSIQKNMVTCHRDKNCI 131

QY 127 INKVTNRCQYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPEVGELIEKV 186

Db 132 INKVTNRCQYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPEVGELIEKV 191

QY 187 RKAHQETFPALCOLGKYTTNNSSEQRVSLDIDLWDFSELSTKCIITVEFAKOLPGFTT 246

Db 192 RKAHQETFPALCOLGKYTTNNSSEQRVSLDIDLWDFSELSTKCIITVEFAKOLPGFTT 251

QY 247 LTIAQDITLLKAAACDILILIRICTRYTPQDQDLEQDPRVDMLOEPLEALKYVYKRRPSRP 306

Db 252 LTIAQDITLLKAAACDILILIRICTRYTPQDQDLEQDPRVDMLOEPLEALKYVYKRRPSRP 311

QY 307 FANQLLPLEMDDAETGLLSAICLICGDRDLEQDPRVDMLOEPLEALKYVYKRRPSRP 366

Db 312 FANQLLPLEMDDAETGLLSAICLICGDRDLEQDPRVDMLOEPLEALKYVYKRRPSRP 371

QY 367 HMFPMKMKITDLRSISAKGAERVITLKMIEPGSMPLIQEMLENSEGLDTLSGQPGGGG 426

Db 372 HMFPMKMKITDLRSISAKGAERVITLKMIEPGSMPLIQEMLENSEGLDTLSGQPGGGG 431

QY 427 RDGGLPPLPPGSCSPSLPSSNRSSPATHSP 457

Db 432 RDGGLAPPPGSCSPSLPSSNRSSPATHSP 462

RESULT 6

AAY21636

ID AAY21636 standard; protein; 416 AA.

XX AC AAY21636;

XX DT 11-AUG-1999 (first entry)

XX DE Ligand binding domain of nuclear receptor hrARalpha.

XX KW Thyroid hormone receptor; aromatic compound; ligand binding domain;

XX KW alpha-glycerophosphate dehydrogenase; cardiac; obesity; triglyceride;

XX KW plasma cholesterol; anti-hypertriglyceridaemic; atherosclerosis; GPDH;

XX KW thyroid hormone replacement therapy; nuclear receptor.

XX OS Homo sapiens.

XX PN W09926966-A2.

XX PD 03-JUN-1999.

XX PF 25-NOV-1998; 98WO-US25296.

XX PR 26-NOV-1997; 97US-0980115.

XX PA (REGC ) UNIV CALIFORNIA.

XX PI Aprilletti JW, Baxter JD, Fletterick RJ, Kushner PJ;

XX PI Scanlan TS, Shiau AK, Wagner RL, West BL;

XX DR WPI; 1999-357810/30.

XX PT Modulating activity of a thyroid hormone receptor





CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. ABG00010-ABG30377 represent novel human  
 CC diagnostic amino acid sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPO  
 CC at ftp.wipo.int/pub/published\_pct\_sequences.

XX SQ Sequence 668 AA;

Query Match 86.7%; Score 2112; DB 22; Length 668;  
 Best Local Similarity 99.3%; Pred. No. 1.1e-172;  
 Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 55 SIETQSSSEIIVPSPPPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKN 114  
 :|||||  
 Db 266 AIETQSSSEIIVPSPPPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKN 325  
 :|||||

QY 115 MYVTCHRDKNKIINKVTRNPCOYCRLOKCFEVMGSKESVRNDRNKKKKEVPKPCSESYT 174  
 :|||||  
 Db 326 MYVTCHRDKNKIINKVTRNPCOYCRLOKCFEVMGSKESVRNDRNKKKKEVPKPCSESYT 385  
 :|||||

QY 175 LPEVGEELIEKVRKAHOETFPALCOLGKYTTNNSSQRYSLDIDLWDFSELSTKCIIT 234  
 :|||||  
 Db 386 LPEVGEELIEKVRKAHOETFPALCOLGKYTTNNSSQRYSLDIDLWDFSELSTKCIIT 445  
 :|||||

QY 235 VEFKQLPGFTTLTIADQITLLKAACLDILILICITRYTPEODTMTFSDGLTLNRTOMHN 294  
 :|||||  
 Db 446 VEFKQLPGFTTLTIADQITLLKAACLDILILICITRYTPEODTMTFSDGLTLNRTOMHN 505  
 :|||||

QY 295 AGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLCGDRDLEQDPRVDMLOEPLEAL 354  
 :|||||  
 Db 506 AGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLCGDRDLEQDPRVDMLOEPLEAL 565  
 :|||||

QY 355 KYVVRKRPSRPHMPKMLKITDLRSISAKGAERVITLTKMEIPGSMPLIQEMLENSEG 414  
 :|||||  
 Db 566 KYVVRKRPSRPHMPKMLKITDLRSISAKGAERVITLTKMEIPGSMPLIQEMLENSEG 625  
 :|||||

QY 415 LDTLSQPGGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457  
 :|||||  
 Db 626 LDTLSQPGGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 668  
 :|||||

RESULT 10  
 AAR27533  
 ID AAR27533 standard; Protein; 797 AA.

XX AC AAR27533;

XX DT 09-MAR-1993 (first entry)

XX DE myl/RAR-alpha fusion.

XX KW Retinoic acid receptor; RAR-alpha; myl; acute promyelocytic leukemia;  
 KW APL; translocation; chromosome 17; chromosome 15; PCR; primer;  
 KW [t(15;17)(q21;q11-22)]; breakpoint; polymerase chain reaction.

XX OS Synthetic.

XX PN W09216660-A.

XX PD 01-OCT-1992.

XX PF 23-MAR-1992; 92WO-US02320.

XX XX 22-MAR-1991; 91US-0673838.

XX PR 22-MAR-1991; 91US-0675084.

XX XX

PA (SLOK ) SLOAN KETTERING INST CANCER.  
 XX Dmitrovsky E, Evans RM, Frankel S, Kazizuka A, Miller WH;  
 PI Warrell RP;  
 XX WPI; 1992-349240/42.  
 DR N-PSDB; AAQ29334.  
 XX Marker for acute promyelocytic leukaemia and other neoplasias  
 PT comprising nucleic acid and encoded abnormal retinoic acid  
 PT receptor-alpha receptor  
 XX Disclosure; Fig 4; 84pp; English.

XX The sequence given shows a fusion between retinoic acid receptor  
 CC (RAR)-alpha and myl which is characteristic of acute promyelocytic  
 CC leukemia (APL). This is caused by a translocation of a portion of  
 CC the long arm of chromosome 17 onto the long arm of chromosome 15  
 CC [t(15;17)(q21;q11-22)]. The breakpoint region has been cloned and it  
 CC has been shown that DNA rearrangements are clustered in the region  
 CC of the first intron of RAR-alpha. This sequence was isolated by  
 CC polymerase chain reaction (PCR) using primers which correspond to  
 CC sequences both 5' and 3' to the breakpoint region.

XX SQ Sequence 797 AA;

Query Match 86.7%; Score 2112; DB 13; Length 797;  
 Best Local Similarity 99.3%; Pred. No. 1.5e-172;  
 Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 55 SIETQSSSEIIVPSPPPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKN 114  
 :|||||  
 Db 395 AIETQSSSEIIVPSPPPPLPRIYKPCFVCQDKSSGYHGVSAECGCKGFFRRSIQKN 454  
 :|||||

QY 115 MYVTCHRDKNKIINKVTRNPCOYCRLOKCFEVMGSKESVRNDRNKKKKEVPKPCSESYT 174  
 :|||||  
 Db 455 MYVTCHRDKNKIINKVTRNPCOYCRLOKCFEVMGSKESVRNDRNKKKKEVPKPCSESYT 514  
 :|||||

QY 175 LPEVGEELIEKVRKAHOETFPALCOLGKYTTNNSSQRYSLDIDLWDFSELSTKCIIT 234  
 :|||||  
 Db 515 LPEVGEELIEKVRKAHOETFPALCOLGKYTTNNSSQRYSLDIDLWDFSELSTKCIIT 574  
 :|||||

QY 235 VEFKQLPGFTTLTIADQITLLKAACLDILILICITRYTPEODTMTFSDGLTLNRTOMHN 294  
 :|||||  
 Db 575 VEFKQLPGFTTLTIADQITLLKAACLDILILICITRYTPEODTMTFSDGLTLNRTOMHN 634  
 :|||||

QY 295 AGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLCGDRDLEQDPRVDMLOEPLEAL 354  
 :|||||  
 Db 635 AGFGPLTDLVFAFANQLPLEMDDAETGLLSAICLCGDRDLEQDPRVDMLOEPLEAL 694  
 :|||||

QY 355 KYVVRKRPSRPHMPKMLKITDLRSISAKGAERVITLTKMEIPGSMPLIQEMLENSEG 414  
 :|||||  
 Db 695 KYVVRKRPSRPHMPKMLKITDLRSISAKGAERVITLTKMEIPGSMPLIQEMLENSEG 754  
 :|||||

QY 415 LDTLSQPGGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457  
 :|||||  
 Db 755 LDTLSQPGGGGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 797  
 :|||||

RESULT 11

AAW81963

ID AAW81963 standard; Protein; 797 AA.

XX AC AAW81963;

XX DT 05-FEB-1999 (first entry)

XX DE Human myl/RAR-alpha fusion protein.

XX KW Fusion protein; myl; retinoic acid receptor-alpha; RAR; human;

XX KW acute promyelocytic leukaemia; APL; t(15;17); translocation;

XX KW treatment; all-trans retinoic acid.

XX XX

OS Homo sapiens.  
XX Synthetic.  
PN US5843642-A.  
XX  
PD 01-DEC-1998.  
XX  
PF 21-JUL-1993; 93US-0095728.  
XX  
PR 21-JUL-1993; 93US-0095728.  
PR 22-MAR-1991; 91US-0673838.  
PR 22-MAR-1991; 91US-0675084.  
XX  
XX (SLOK ) SLOAN KETTERING INST CANCER RES.  
XX  
XX Dmirovsky E, Frankel S, Miller WH, Warrell RP;  
XX WPI; 1999-044563/04.  
DR N-PSDB; AAV64990.  
XX  
XX diagnosis of acute promyelocytic leukaemia - by detecting nucleic  
PT acid encoding abnormal retinoic acid receptor-alpha  
XX  
PS Disclosure; Fig 4A-G; 38pp; English.  
XX  
XX This sequence represents a human myl/RAR-alpha (retinoic acid receptor  
CC alpha) fusion protein which is used in a method for identifying a  
CC subject with acute promyelocytic leukaemia (APL) resulting from a  
CC t(15;17) translocation who will respond to treatment with all-trans  
CC retinoic acid. The protein can also be used to identify a subject  
CC with indications of APL who will not respond to treatment with all-trans  
CC retinoic acid.  
XX  
XX Sequence 797 AA;  
XX  
Query Match 86.7%; Score 2112; DB 20; Length 797;  
Best Local Similarity 99.3%; Pred. No. 1.5e-172;  
Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 55 SIETQSSSEIEIVPSPPLPRYKPCFVQCDKSSGYHYGSACGCKGFFRRSIQKN 114  
Db 395 AIETQSSSEIEIVPSPPLPRYKPCFVQCDKSSGYHYGSACGCKGFFRRSIQKN 454  
QY 115 MYTCHRDNCIINKVTRNCPQYCRLOKCFEVMGSKESVRNDRNKKKKEVPKPCSESYT 174  
Db 455 MYTCHRDNCIINKVTRNCPQYCRLOKCFEVMGSKESVRNDRNKKKKEVPKPCSESYT 514  
QY 175 LTPVEGELIEKVRKAHQETFPALCOLGKYTTNNSSQYRVSLLDLDLWDFSELSTKCIKT 234  
Db 515 LTPVEGELIEKVRKAHQETFPALCOLGKYTTNNSSQYRVSLLDLDLWDFSELSTKCIKT 574  
QY 235 VEFKQLPGFTLTADQITLLKAACLDLILIRICTRYTPEQDTWTFSDGLTLNRTOMHN 294  
Db 575 VEFKQLPGFTLTADQITLLKAACLDLILIRICTRYTPEQDTWTFSDGLTLNRTOMHN 634  
QY 295 AGFGLTDLVFAFANQLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEAL 354  
Db 635 AGFGLTDLVFAFANQLPLEMDDAETGLLSAICLICGDRQDLEQPDVMDLQEPLEAL 694  
QY 355 KYVVKRRPSRPHMPKMLMKITDLRSISAKGAERVITLKMIEIPGSMPLIOBMLENSEG 414  
Db 695 KYVVKRRPSRPHMPKMLMKITDLRSISAKGAERVITLKMIEIPGSMPLIOBMLENSEG 754  
QY 415 LDTLSGQPGGGGDDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457  
Db 755 LDTLSGQPGGGGDDGGGLPPPGSCSPSLSPSSNRSSPATHSP 797  
RESULT 12  
ABG18860  
ID ABG18860 standard; Protein; 819 AA.  
XX  
AC ABG18860;

XX 18-FEB-2002 (first entry)  
DT Novel human diagnostic protein #18851.  
XX  
DE Human; chromosome mapping; gene mapping; gene therapy; forensic;  
XX food supplement; medical imaging; diagnostic; genetic disorder.  
KW  
KW Homo sapiens.  
XX  
OS WO200175067-A2.  
XX  
PN 11-OCT-2001.  
XX  
PD 30-MAR-2001; 2001WO-US08631.  
XX  
PF 31-MAR-2000; 2000US-0540217.  
XX  
PR 23-AUG-2000; 2000US-0649167.  
PR  
XX (HYSE-) HYSEQ INC.  
PA Drmanac RT, Liu C, Tang YT;  
XX WPI; 2001-639362/73.  
PI N-PSDB; AAS83047.  
XX  
DR New isolated polynucleotide and encoded polypeptides, useful in  
XX diagnostics, forensics, gene mapping, identification of mutations  
XX responsible for genetic disorders or other traits and to assess  
XX biodiversity  
XX Claim 20; SEQ ID NO 49219; 103pp; English.  
XX  
XX The invention relates to isolated polynucleotide (I) and  
CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
CC and gene mapping, and in recombinant production of (II). The  
CC polynucleotides are also used in diagnostics as expressed sequence tags  
CC for identifying expressed genes. (I) is useful in gene therapy techniques  
CC to restore normal activity of (II) or to treat disease states involving  
CC (II). (II) is useful for generating antibodies against it, detecting or  
CC quantitating a polypeptide in tissue, as molecular weight markers and as  
CC a food supplement. (II) and its binding partners are useful in medical  
CC imaging of sites expressing (II). (I) and (II) are useful for treating  
CC disorders involving aberrant protein expression or biological activity.  
CC The polypeptide and polynucleotide sequences have applications in  
CC diagnostics, forensics, gene mapping, identification of mutations  
CC and to produce other types of data and products dependent on DNA and  
CC amino acid sequences. ABG00010-ABG30377 represent novel human  
CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed  
CC specification, but was obtained in electronic format directly from WIPO  
CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
XX Sequence 819 AA;

Query Match 86.7%; Score 2112; DB 22; Length 819;  
Best Local Similarity 99.3%; Pred. No. 1.5e-172;  
Matches 400; Conservative 1; Mismatches 2; Indels 0; Gaps 0;  
QY 55 SIETQSSSEIEIVPSPPLPRYKPCFVQCDKSSGYHYGSACGCKGFFRRSIQKN 114  
Db 417 AIETQSSSEIEIVPSPPLPRYKPCFVQCDKSSGYHYGSACGCKGFFRRSIQKN 476  
QY 115 MYTCHRDNCIINKVTRNCPQYCRLOKCFEVMGSKESVRNDRNKKKKEVPKPCSESYT 174  
Db 477 MYTCHRDNCIINKVTRNCPQYCRLOKCFEVMGSKESVRNDRNKKKKEVPKPCSESYT 536  
QY 175 LTPVEGELIEKVRKAHQETFPALCOLGKYTTNNSSQYRVSLLDLDLWDFSELSTKCIKT 234  
Db 537 LTPVEGELIEKVRKAHQETFPALCOLGKYTTNNSSQYRVSLLDLDLWDFSELSTKCIKT 596







```
Query Match      83.2%; Score 2026.5; DB 22; Length 470;
Best Local Similarity 88.5%; Pred. No. 1.6e-165;
Matches 402; Conservative 7; Mismatches 22; Indels 23; Gaps 7;

QY 7 VGGTNPFLVVDYFNQNRACLLPERGLPAGPYSTPLRPLWNGSNHSTETOSSEEEI 66
Db 37 LGGLSP-PCALTTLQHQ-----LPVSG-YSTP-----SPATETOSSEEEI 76

QY 67 VSPSPSPPLPRIYKPCFYCDKSSGYHYGSACGCKGFFRRSIQKNMYYTCHRDKNCI 126
Db 77 VSPSPSPPLPRIYKPCFYCDKSSGYHYGSACGCKGFFRRSIQKNMYYTCHRDKNCI 136

QY 127 INKVTNPQCYCRLOKCFEYVGMKESVRNDRNKKKEVPKPECSSESYTLPEVGELIEKV 186
Db 137 INKVTNRQCVCRLQKCFEYVGMKESVRNDRNKKKEVPKPECSSESYTLPEVGELIEKV 196

QY 187 RKAHQETFFALCOLGKYTTNNSSEQRVSLDIDLWDKFSELSTKCIITKTVFEAK-OLPGFT 245
Db 197 RKAHQETFFALCOLGKYTTNNSSEQRVSLDIDLWDKFSELSTKCIITKNGEGROGLPGFT 256

QY 246 TLTIAQITILLKAACLDILILRICTRYTPEQDTMTFS-DGLTLNRTOMHNAGFGLTDLV 304
Db 257 TLTIAQITILLKAACLDILILRICTRYTPEAGTPXPFSWDGLTLNRTOMHNAGFGLTDLV 316

QY 305 FAFANQLLPLEMDDA-ETGLLSAICLCGRDQDLEQPDVDMLEQPLLEALKVYVRKRRP 363
Db 317 FAFANQLLPLEMDDAETGLLSAICLCGRDQDLEQPDVDMLEQPLLEALKVYVRKRRP 376

QY 364 SRPHMFPKMLKITDLRSISAKGAERVITLKMPIGSMPLIOEMLENSEGLDTLSCQPG 423
Db 377 SRPHMFPKMLKITDLRSISAKGAERVITLKMPIGSMPLIOEMLENSEGLDTLSCQPG 436

QY 424 GGRDGGGLPPPGSCSPSLSPSSNRSSPATHSP 457
Db 437 GGRDGGGLAPPGGSCSPSLSPSSNRSSPATHSP 470
```

Search completed: March 29, 2003, 06:38:55  
Job time : 58 secs

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	2127	87.3	462	10	US-09-797-727-3		Sequence 3, Appli
2	2124	87.2	462	9	US-09-814-604-1		Sequence 1, Appli
3	1771	72.7	448	9	US-09-814-604-2		Sequence 2, Appli
4	1765	72.4	448	10	US-09-797-727-4		Sequence 4, Appli
5	1662	68.2	454	9	US-09-814-604-3		Sequence 3, Appli
6	1662	68.2	454	10	US-09-797-727-2		Sequence 2, Appli
7	560	23.0	461	12	US-10-013-823-3		Sequence 3, Appli
8	549	22.5	446	10	US-09-909-448-2		Sequence 2, Appli
9	549	22.5	446	10	US-09-909-328-2		Sequence 2, Appli
10	549	22.5	446	10	US-09-909-328-2		Sequence 2, Appli
11	545	22.4	446	12	US-10-013-823-2		Sequence 2, Appli
12	519.5	21.3	598	10	US-09-853-386-32		Sequence 32, Appl
13	519.5	21.3	598	10	US-09-853-386-35		Sequence 35, Appl
14	519.5	21.3	598	10	US-09-853-386-36		Sequence 36, Appl
15	519.5	21.3	598	10	US-09-853-386-37		Sequence 37, Appl
16	519.5	21.3	598	10	US-09-853-386-38		Sequence 38, Appl
17	519.5	21.3	598	10	US-09-853-386-39		Sequence 39, Appl
18	519.5	21.3	598	10	US-09-853-386-40		Sequence 40, Appl
19	519.5	21.3	598	10	US-09-853-386-41		Sequence 41, Appl



```
; TITLE OF INVENTION: GAMMA RETINOIC ACID RECEPTOR
; FILE REFERENCE: SALK1150-3
; CURRENT APPLICATION NUMBER: US/09/797,727
; CURRENT FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 08/486,325
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/100,039
; PRIOR FILING DATE: 1993-07-30
; PRIOR APPLICATION NUMBER: PCT/US90/03564
; PRIOR FILING DATE: 1990-06-22
; PRIOR APPLICATION NUMBER: US 07/370,407
; PRIOR FILING DATE: 1989-06-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 4
; LENGTH: 448
; TYPE: PRT
; ORGANISM: Human
; NAME/KEY: misc-feature
; OTHER INFORMATION: Human Retinoic Acid Receptor-beta (hRAR-beta)
US-09-797-727-4
```

```
Query Match 72.4%; Score 1765; DB 10; Length 448;
Best Local Similarity 74.1%; Pred. No. 2.8e-136;
Matches 341; Conservative 43; Mismatches 58; Indels 18; Gaps 5;

QY 1 MYSEVEGGPTNPFVVDYFNON--RACLLPEKGLPAPGPTSTPLRTPLWNGSN--HSIE 57
Db 1 MFDGMDVLSVSPGQ--ILDFTASPSOCLQEKALKA--CFSGLTQTWQHRHTAQSTIE 55

QY 58 TQSSSEIEVPSPPPPLPRIYPCFCVQDKSSGHHYGVSAECGCKGFFRRSIQKNVY 117
Db 56 TQTSSEELVSPSPPLPPRPVYPCFCVQDKSSGHHYGVSAECGCKGFFRRSIQKNVY 115

QY 118 TCHRDKNCKVINKVTRNCOYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTTP 177
Db 116 TCHRDKNCKVINKVTRNCOYCRLOKCFEVMGSKESVRNDRNKKKEVSKQECTSEYMTA 175

QY 178 EVGLIEKVRKAHQETPPALCQLGKYTTNNSSEQRVSLDIDLMDKFSSELTKCIKTVEF 237
Db 176 ELDDLTEKIRKAHQETPPSLCQLGKYTTNNSADHRVRLDLGLMDKFSSELATKCIKIVEF 235

QY 238 AKQLPGFTTLTADQITLLKAAACDILILRICTRYTPQDTMTFSDGLTLNRTQHNAGF 297
Db 236 AKRLPGFTGLSIADQITLLKAAACDILILRICTRYTPQDTMTFSDGLTLNRTQHNAGF 295

QY 298 GPLDLVFAPANQLPLMDMDAETGLLSAICLCGDRDLQEPDRVDMQLQEPPLLEALKVY 357
Db 296 GPLDLVTFANQLPLMDMDTETGLLSAICLCGDRDLQEPDRVDMQLQEPPLLEALKIY 355

QY 358 VRKRPRSRPHFPKMLKITDLRSISAKGAERVITLKMIEPGSMPLLIQEMLENSEGLDT 417
Db 356 IRKRPRSPHFPKMLKITDLRSISAKGAERVITLKMIEPGSMPLLIQEMLENSEGLDT 415

QY 418 LSGPGGGGRDGGGLPPPGSCSLSLSPSSNRKSPATHSP 457
Db 416 LTPSSSGNTAEH-----SPSISPSSVNSGVSGSQSP 445

RESULT 5
US-09-814-604-3
; Sequence 3, Application US/09814604
; Publication No. US20030003517A1
; GENERAL INFORMATION:
; APPLICANT: Klein, Elliott S.
; TITLE OF INVENTION: Methods of Detecting Dissociated Nuclear
; TITLE OF INVENTION: Hormone Receptor Ligands
; FILE REFERENCE: P-AR 4528
; CURRENT APPLICATION NUMBER: US/09/814,604
; CURRENT FILING DATE: 2001-03-22
; NUMBER OF SEQ ID NOS: 52
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-814-604-3

Query Match 68.2%; Score 1662; DB 9; Length 454;
Best Local Similarity 72.8%; Pred. No. 7.5e-128;
Matches 319; Conservative 39; Mismatches 48; Indels 32; Gaps 4;

QY 30 PEKGLPAPG-PYSTP-----LRTPLWNG-----SNHSIETQSSSEIIVP 68
Db 16 PGSGYPGAGFPFAFGALRGSPFFEMLSPSFRGLGQDLPKEMASLSVETQSTSSSEMV 75

QY 69 SPPSPPLPRIYKPCFVQDKSSGHHYGVSAECGCKGFFRRSIQKNVYTCRDKNCIIN 128
Db 76 SSPSPPLPRIYKPCFVQDKSSGHHYGVSAECGCKGFFRRSIQKNVYTCRDKNCIIN 135

QY 129 KYTRNCPQYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPVGGELIEKVRK 188
Db 136 KYTRNCPQYCRLOKCFEVMGSKESVRNDRNKKKEVPKPCSESYTLTPVGGELIEKVRK 195

QY 189 AHQETFPALCQLGKYTTNNSSEQRVSLDIDLMDKFSSELTKCIKTVEFAKQLPGFTTTL 248
Db 196 AHQETFPALCQLGKYTTNNSADHRVRLDLGLMDKFSSELATKCIKTVEFAKRLPGFTGLS 255

QY 249 IADQITLLKAAACDILILRICTRYTPQDTMTFSDGLTLNRTQHNAGFGLTLDLVFAFA 308
Db 256 IADQITLLKAAACDILILRICTRYTPQDTMTFSDGLTLNRTQHNAGFGLTLDLVFAFA 315

QY 309 NOLLPLEMDDAETGLLSAICLCGDRDLQEPDRVDMQLQEPPLLEALKVYVYKRRPSRPHM 368
Db 316 GOLLPLEMDDTETGLLSAICLCGDRDLQEPDRVDMQLQEPPLLEALKVYVYKRRPSQPYM 375

QY 369 FPKMLKITDLRSISAKGAERVITLKMIEPGSMPLLIQEMLENSEGLDTLSGGPG----- 423
Db 376 FPKMLKITDLRSISAKGAERVITLKMIEPGSMPLLIQEMLENSEGLDTLSGGPGPHNA 435

QY 424 -----GGGRDGGGLPPP 435
Db 436 SSEDEVPGGGGGLKSP 453

RESULT 6
US-09-797-727-2
; Sequence 2, Application US/09797727
; Patent No. US20020077457A1
; GENERAL INFORMATION:
; APPLICANT: the Salk Institute for Biological Studies
; APPLICANT: TAKAKU, Fumimaro
; TITLE OF INVENTION: GAMMA RETINOIC ACID RECEPTOR
; FILE REFERENCE: SALK1150-3
; CURRENT APPLICATION NUMBER: US/09/797,727
; CURRENT FILING DATE: 2001-08-31
; PRIOR APPLICATION NUMBER: US 08/486,325
; PRIOR FILING DATE: 1995-06-07
; PRIOR APPLICATION NUMBER: US 08/100,039
; PRIOR FILING DATE: 1993-07-30
; PRIOR APPLICATION NUMBER: PCT/US90/03564
; PRIOR FILING DATE: 1990-06-22
; PRIOR APPLICATION NUMBER: US 07/370,407
; PRIOR FILING DATE: 1989-06-22
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 2
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Human
; NAME/KEY: misc-feature
; OTHER INFORMATION: Human Retinoic Acid Receptor-gamma (hRAR-gamma)
US-09-797-727-2
```

Query Match 68.2%; Score 1662; DB 10; Length 454;  
Best Local Similarity 72.8%; Pred. No. 7.5e-128;  
Matches 319; Conservative 39; Mismatches 48; Indels 32; Gaps 4;

QY 30 PEKGLPAG-PYSTP-----LRTPLWNG-----SNHSIQTSSSEIIVP 68  
DB 16 PGSGYPCAGFPFAFGALRGSPFEMLSFSGRLGQPDLPKEMASLVSFQTSSEEMVP 75

QY 69 SPSPPLPRYKPCFCVQCKSSGYHYGSACGCKGFFRRSIOKNNVYTCCHRDKNCIIN 128  
DB 76 SPSPPLPRYKPCFCVQCKSSGYHYGSACGCKGFFRRSIOKNNVYTCCHRDKNCIIN 135

QY 129 KVTNRNCOYCRLOKQCEVGMKESVRNDRNKKKEVPKPECSESYTLTPVGELIEKVRK 188  
DB 136 KVTNRNCOYCRLOKQCEVGMKESVRNDRNKKKEVPKPECSESYTLTPVGELIEKVRK 195

QY 189 AHOETFPALCOLGKYTYNNSSEORVSLDLDLWDFKSELSTKCIIVKVEFAKQLPGFTTTL 248  
DB 196 AHOETFPALCOLGKYTYNNSSEORVSLDLDLWDFKSELSTKCIIVKVEFAKQLPGFTTTL 255

QY 249 IADQITLLAAACILILICRITRYTEQDTMTFSDGLTLNRTQHNAGFGLDLYFAFA 308  
DB 256 IADQITLLAAACILILICRITRYTEQDTMTFSDGLTLNRTQHNAGFGLDLYFAFA 315

QY 309 NOLLPLEMDAETGLLSAICLCGRDQLEQDPRVDMQLQEPLEALKVYVYKRRRPSRPHM 368  
DB 316 GOLLPLEMDTETGLLSAICLCGRDQLEQDPRVDMQLQEPLEALKVYVYKRRRPSRPHM 375

QY 369 FPKMLMKITDLSISAKGAERVITLMEIPGSMPLIQEMLNSEGDLTSLGQFG----- 423  
DB 376 FPKMLMKITDLSISAKGAERVITLMEIPGSMPLIQEMLNSEGDLTSLGQFG----- 435

QY 424 -----GGGRDGGGLPPP 435  
DB 436 SSEDEVPGGQGGGLKSP 453

RESULT 7  
US-10-013-823-3  
; Sequence 3, Application US/10013823  
; Patent No. US20020116731A1  
; GENERAL INFORMATION:  
; APPLICANT: Guenther, Catherine  
; APPLICANT: Phillips, Russell  
; APPLICANT: Allen, Keith D.  
; APPLICANT: Zhang, Qin  
; APPLICANT: Baribault, Helene  
; TITLE OF INVENTION: TRANSGENIC MICE CONTAINING RETINOID X  
; FILE REFERENCE: R-684  
; CURRENT FILING DATE: 2001-12-10  
; PRIOR APPLICATION NUMBER: US 60/254,801  
; PRIOR FILING DATE: 2000-12-11  
; PRIOR APPLICATION NUMBER: US 60/309,404  
; PRIOR FILING DATE: 2001-07-31  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 461  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-013-823-3

Query Match 23.0%; Score 560; DB 12; Length 461;  
Best Local Similarity 30.6%; Pred. No. 8.8e-38;  
Matches 140; Conservative 82; Mismatches 155; Indels 80; Gaps 12;

QY 29 LPEKGLPAG-PYSTPL-----RTPLWNGSNHSI-----ETOSSESSEIVPSPPP--- 74  
DB 13 LFGNGPQPGAPSSPTVKKEGPEWPGPDVPGTDASSACSTDWVDPPEEPERK 72

QY 75 ----PLPRI--YKPCFVCQDKSSGYHYGSACGCKGFFRBSIOKNNV--YTCHRDKNCI 126  
DB 73 RKGPAPMGLGHELCRCVCGDKASGFHYNVLSCEGCKGFFRVSVVRGGARRACRGCGTCQ 132

QY 127 INKVTNRNCOYCRLOKQCEVGMKESVRNDRNKKKEVPKPECSESYTLT----- 176  
DB 133 MDAPMRKCKQOCLRKCKEAGMRQCVLSEBQIRKKKIRKQOQESQSQSPVGPQGS 192

QY 177 -----PVVG-----ELIEKVRKAHQTFPFCALQGLKYT----- 204  
DB 193 SSASGPGASPGSGSAGSGGSEGVQLTAAQELMIQOLVAAQCNKRFSQDPKVTWP 252

QY 205 -----TNSSEORVSLDLDLWDFKSELSTKCIIVKVEFAKQLPGFTTTLTADQITLLK 257  
DB 253 PLGADPOSRAQRQFA-----HFTELAIISVQEIIVDFAKQVPGFLOLQREGDQIALLK 305

QY 258 AACDLILILICRITRYTEQDTMTFSDGLTLNRTQHNAGFG-PLTDLVFAFANQLLPLEM 316  
DB 306 ASTIEIMLLETARRYNHETECITFLKDTYSKDDFHRAGLOVEFINPIFEFSRAMRRLGL 365

QY 317 DDAETGLLSAICLCGRDQLEQDPRVDMQLQEPLEALKVYVYKRRRPSRPHMFKMLMKI 376  
DB 366 DDAEYALIIAINIFSADRPNVQPRVEALQOQPYVEALLSYTRIKRPODQRFPRMLMKL 425

QY 377 TDLRSISAKGAERVITLMEIPGSMPLIQEMLNSEGDLTSLGQFG----- 413  
DB 426 VSLRTSLSSVHSEQVFAIRLQ-DKKLPPLLSEIWDVHE 461

RESULT 8  
US-09-909-446-2  
; Sequence 2, Application US/09909446  
; Patent No. US20020052489A1  
; GENERAL INFORMATION:  
; APPLICANT: ENMARK, EVA  
; GUSTAFSSON, JAN  
; TITLE OF INVENTION: OR-1 ON ORPHAN RECEPTOR BELONGING  
; TO THE NUCLEAR RECEPTOR FAMILY  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Banner & Witcoff  
; STREET: 1001 G Street, NW  
; CITY: Washington  
; STATE: DC  
; COUNTRY: USA  
; ZIP: 20001  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Diskette  
; COMPUTER: IBM Compatible  
; OPERATING SYSTEM: DOS  
; SOFTWARE: FastSeq for Windows Version 2.0  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/909,446  
; FILING DATE: 19-Jul-2001  
; CLASSIFICATION: <Unknown>  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/776,844  
; FILING DATE: <Unknown>  
; APPLICATION NUMBER: UK 9413536.2  
; FILING DATE: 16-AUG-1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Kegan, Sarah A  
; REGISTRATION NUMBER: 32141  
; REFERENCE/DOCKET NUMBER: 00487.04029  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-508-9100  
; TELEFAX: 202-508-9299  
; TELEX: <Unknown>  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 446 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single



```
;
;
; MOLECULE TYPE: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-909-446-2

Query Match      22.5%; Score 549; DB 10; Length 446;
Best Local Similarity 31.6%; Pred. No. 6.7e-37;
Matches 141; Conservative 75; Mismatches 158; Indels 72; Gaps 12;

QY 29 LPEKGLPAPGYPSTPLRTPLWNGSNHSIE-----TQSSSSEIYP--SPSPSP-----P 75
Db 12 LPNGSPQPSSTST---SPTIKEGQETDPPPGSEGSSAYIVVILEPEDEPERKRRKGP 68

QY 76 LPRI--YKCFVCODKSSGYHYGSACEGCKGFFRRSIQKNV--YTCRDKNCIINKVT 131
Db 69 APRMLGHELRCVCGDKASGFHYNVLSCEGCKGFFRRSVVHGGAGRYACRGSGTCOMDAFM 128

QY 132 RNPQCYCRLOKCFEYGMKSERYNDRNKKKEVPK-----PE 168
Db 129 RKKQLCRLRKCKEAGMREQCVLSEEQIRKKKIQAQLOCNKRSFSDQPKVTPWPLGADPQSRDA 188

QY 169 CSSEYTLTPEVGELEIKVRAH---QETFPALCOLGKYT-----TNNS 208
Db 189 TSEASSQSGEGEGIQLTAAQELMIQQLVAQLOCNKRSFSDQPKVTPWPLGADPQSRDA 248

QY 209 SEQVSLDIDLWDKFSLSKCIKTVEFAKQLPGFTTITADQITLLKAACLDILIRI 268
Db 249 RQOREA-----HFTELAIISVQEIYVDEFAKQVPGFLQGLREDQIALLKASTIEIMLET 301

QY 269 CTRYTEQDWTFSDDLNTQTMHNAFG--PLTDLVFAFANQLLPLEMDDAETGLLSAI 327
Db 302 ARRYNHETECITFLKDFTSKDDPHRAGLOVEFINPIFEFSRAMRRLGLDDAEYALLIAI 361

QY 328 CLICGRDQLEQPDVDMLOEPLEALKVYVRRKRRSPRPHMFKMLMKITDLRSISAKGA 387
Db 362 NIFSADRPNVQEPSRVEALQOQVVEALLSYTRIKRPQDQLRPPRLMKLVSLRTLSVHS 421

QY 388 ERVITLKMIEIPGSMPLIOEMLENSE 413
Db 422 EQVFALRLQ-DKKLPLLSEIWDVHE 446

RESULT 9
US-09-909-325-2
; Sequence 2, Application US/09909325
; Patent No. US20020115847A1
; GENERAL INFORMATION:
; APPLICANT: ENMARK, EVA
; TITLE OF INVENTION: OR-1 ON ORPHAN RECEPTOR BELONGING
; TO THE NUCLEAR RECEPTOR FAMILY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/909,325
; FILING DATE: 19-Jul-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/776,844
; FILING DATE: <Unknown>
; APPLICATION NUMBER: UK 9413536.2
; FILING DATE: 16-AUG-1994
```

```
;
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A
; REGISTRATION NUMBER: 32141
; REFERENCE/DOCKET NUMBER: 00487.04029
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: <Unknown>
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 446 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-909-325-2

Query Match      22.5%; Score 549; DB 10; Length 446;
Best Local Similarity 31.6%; Pred. No. 6.7e-37;
Matches 141; Conservative 75; Mismatches 158; Indels 72; Gaps 12;

QY 29 LPEKGLPAPGYPSTPLRTPLWNGSNHSIE-----TQSSSSEIYP--SPSPSP-----P 75
Db 12 LPNGSPQPSSTST---SPTIKEGQETDPPPGSEGSSAYIVVILEPEDEPERKRRKGP 68

QY 76 LPRI--YKCFVCODKSSGYHYGSACEGCKGFFRRSIQKNV--YTCRDKNCIINKVT 131
Db 69 APRMLGHELRCVCGDKASGFHYNVLSCEGCKGFFRRSVVHGGAGRYACRGSGTCOMDAFM 128

QY 132 RNPQCYCRLOKCFEYGMKSERYNDRNKKKEVPK-----PE 168
Db 129 RKKQLCRLRKCKEAGMREQCVLSEEQIRKKKIQAQLOCNKRSFSDQPKVTPWPLGADPQSRDA 188

QY 169 CSSEYTLTPEVGELEIKVRAH---QETFPALCOLGKYT-----TNNS 208
Db 189 TSEASSQSGEGEGIQLTAAQELMIQQLVAQLOCNKRSFSDQPKVTPWPLGADPQSRDA 248

QY 209 SEQVSLDIDLWDKFSLSKCIKTVEFAKQLPGFTTITADQITLLKAACLDILIRI 268
Db 249 RQOREA-----HFTELAIISVQEIYVDEFAKQVPGFLQGLREDQIALLKASTIEIMLET 301

QY 269 CTRYTEQDWTFSDDLNTQTMHNAFG--PLTDLVFAFANQLLPLEMDDAETGLLSAI 327
Db 302 ARRYNHETECITFLKDFTSKDDPHRAGLOVEFINPIFEFSRAMRRLGLDDAEYALLIAI 361

QY 328 CLICGRDQLEQPDVDMLOEPLEALKVYVRRKRRSPRPHMFKMLMKITDLRSISAKGA 387
Db 362 NIFSADRPNVQEPSRVEALQOQVVEALLSYTRIKRPQDQLRPPRLMKLVSLRTLSVHS 421

QY 388 ERVITLKMIEIPGSMPLIOEMLENSE 413
Db 422 EQVFALRLQ-DKKLPLLSEIWDVHE 446
```

```
RESULT 10
US-09-909-326-2
; Sequence 2, Application US/09909326
; Patent No. US20020128461A1
; GENERAL INFORMATION:
; APPLICANT: ENMARK, EVA
; GUSTAFSSON, JAN
; TITLE OF INVENTION: OR-1 ON ORPHAN RECEPTOR BELONGING
; TO THE NUCLEAR RECEPTOR FAMILY
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20001
; COMPUTER READABLE FORM:
```



```

; PRIOR APPLICATION NUMBER: US 60/203645
; PRIOR FILING DATE: 2000-03-12
; NUMBER OF SEQ ID NOS: 153
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 32
; LENGTH: 598
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-853-386-32

```

Query Match	21.3%;	Score 519.5;	DB 10;	Length 598;
Best Local Similarity	31.5%;	Pred. NO. 2.6e-34;		
Matches 141;	Conservative	79;	Mismatches 154;	Indels 73;
			Gaps	16;

Query Match	21.3%;	Score 519.5;	DB 10;	Length 598;
Best Local Similarity	31.5%;	Pred. No. 2.6e-34;		

```

Query Match      21.3%; Score 519.5; DB 10; Length 598;
Best Local Similarity 31.5%; Pred. No. 2.6e-34;
Matches 141; Conservative 79; Mismatches 154; Indels 73; Gaps 16;

QY 24 NRACLLPEKGLPAGPYST-----PLRTPLWN-----GSNHSIETQSSSEIIVSP-- 70
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 171 SRLSLFSEKQSPCTPVSSQCMRFDGLHVPN-NPEPAGSHHVVDGTFA----VNPPIR 225
      : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 71 -----PSPPP--LPRIYKPCFVCODKSGSYHYGVSAACEGKG 105
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | |

Db 226 KPASMGFFGLIGHASOLLDTQVPSPPSRCSNPEGLCVCGDNAACQHVGVITCECGKG 285
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 106 FFRRSIOKNMYYTCHRDKNCIINKVTRNCPQQLQKCFEVMGMSKESVRNDRNK-KKKEV 164
      |||| | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Db 286 FFKRTVQKNAKYVCLANKNCVDRKRRNRQCYCRFOKCLAVGMVKEVVRTDLSLKGRRGL 345  
QY 165 P-KPECSSEYLTTPVEGELIEKVRKAHQETFPALCOLG-----KYTTNNSSEQRVSL 215  
Db 346 PSKPKSPQDPSPSPVSLISALVRAHVDSNPAMTSLDYSRFOANPDYQMSGDDTQHI-- 403  
QY 216 DIDLWDFSELSTKCIKIVKFAKQPLPGFTTLTADQITLLKAACLDILILRICTRYTPE 275  
Db 404 -----QQFYDILLTGSMEIIRGWAKEIPGADLPKADQDILLFESAFLELVRLAYRNPV 458  
QY 276 QDTMTFSDGLTNRTQMHNAGCPDLVLVAFANQLLPLEMDDAETGLLSAICLCGRQ 335  
Db 459 EGKLFPCNGVVLHRLQCVR-GFGEWIDSIVFESSNLQNMNIDISAFSCIAALAMVT-ERH 516  
QY 336 DLEQPDVMDLQEPLELLEALKVYVKKRRP--SRPHMFPKMLKITDLRSISAKGAERVITL 393  
Db 517 GLKEPRVEELQNKIVNCLKDHTVFNNGGLNRPNYLSKLLGKLPDLRTCTOGLQRIFYL 576  
QY 394 KME--IPGSMPPLIQEMLENSEGLDYL 418  
Db 577 KLEDLVP---PPAIDKL-----FLDTL 596

## RESULT 15

US-09-853-386-37.

; Sequence 37, Application US/09853386

; Patent No. US20020049151A1

; GENERAL INFORMATION:

; APPLICANT: Murphy, Evelyn

; APPLICANT: Bresnahan, Barry

; APPLICANT: Conneely, Orla

; APPLICANT: Fitzgerald, Oliver

; TITLE OF INVENTION: Therapeutic Approaches to Diseases by Suppression of the NURR

; FILE REFERENCE: P01972051

; CURRENT APPLICATION NUMBER: US/09/853, 386

; PRIOR FILING DATE: 2001-05-11

; PRIOR APPLICATION NUMBER: US 60/203645

; NUMBER OF SEQ ID NOS: 153

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 37

; LENGTH: 598

; TYPE: PRT

; ORGANISM: HUMAN

US-09-853-386-37

Query Match 21.3%; Score 519.5; DB 10; Length 598;  
Best Local Similarity 31.5%; Pred. No. 2.6e-34;  
Matches 141; Conservative 79; Mismatches 154; Indels 73; Gaps 16;

QY 24 NRACLLPEKGLPAGPYST-----PLTPLWN-----GSNHSIETQSSSSEIIVPSP-- 70  
Db 171 SRLSLFSFKQSPGPTPVSSCOMRFDCGLHVPN-NPEPAGSHVVDGQTPA----VNPPIR 225  
QY 71 -----PSPPP--LPRIYKPCFCODKSSGYHYGVYACGCKG 105  
Db 226 KPASMGFPGLQIGHASQLDVTQVPFPPSPSGSNEGLCAVCGDNACQHYGVYRTCEGCKG 285  
QY 106 FFRSIOKNMVTCHRDKNCIINKVTRNCPQYCRKQCFEVMGSKESVRNDRNK-KKKEV 164  
Db 286 FFKRTVQKNAKYVCLANKNCVDRKRRNRQCYCRFOKCLAVGMVKEVVRTDLSLKGRRGL 345  
QY 165 P-KPECSSEYLTTPVEGELIEKVRKAHQETFPALCOLG-----KYTTNNSSEQRVSL 215  
Db 346 PSKPKSPQDPSPSPVSLISALVRAHVDSNPAMTSLDYSRFOANPDYQMSGDDTQHI-- 403  
QY 216 DIDLWDFSELSTKCIKIVKFAKQPLPGFTTLTADQITLLKAACLDILILRICTRYTPE 275  
Db 404 -----QQFYDILLTGSMEIIRGWAKEIPGADLPKADQDILLFESAFLELVRLAYRNPV 458  
QY 276 QDTMTFSDGLTNRTQMHNAGCPDLVLVAFANQLLPLEMDDAETGLLSAICLCGRQ 335

GenCore version 5.1.4\_p5\_4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 30, 2003, 08:00:38 ; Search time 2120.41 Seconds  
(without alignments)  
13725.086 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_1\_1000  
Perfect score: 1000  
Sequence: 1 gtccttggttagcatgtaca.....acatgcgcgcctctccctctc 1000

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pi.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	973.6	97.4	16913	9	HSA297538	AJ297538 Homo sapi
C 2	973.6	97.4	138999	9	AC090426	Homo sapi
C 3	973.6	97.4	173441	2	AC080112	Homo sapi
C 4	962.6	96.3	188574	2	AC131063	Homo sapi
C 5	962.6	96.3	190309	2	AC126392	Homo sapi
C 6	956.8	95.7	158766	2	AC015851	Homo sapi
7	875.6	87.6	1833	9	AF283809	Homo sapi
8	263.2	26.3	231248	2	AL591067	Mus muscu
9	259	25.9	1330	10	MUSRA21	Mouse reti
C 10	242.8	24.3	13508	6	AX344791	Sequence
C 11	228	22.8	2008	9	AK098172	Homo sapi
C 12	143.8	14.4	13508	6	AX344790	Sequence
C 13	127	12.7	162393	2	AC111746	Rattus no
C 14	73.8	7.4	972	11	PM12A12G	Penicilli
C 15	72.6	7.3	7218	6	I66494	Sequence 14
C 16	69.4	6.9	90352	2	AC099278	Rattus no
C 17	69.2	6.9	207420	2	AC078884	Mus muscu
C 18	68.8	6.9	146395	2	AC127918	Rattus no
C 19	67.4	6.7	300695	2	AC079431	Mus muscu
C 20	67	6.7	94029	2	AC117309	Rattus no
C 21	66.6	6.7	152261	2	AC121010	Rattus no
C 22	66.4	6.6	112299	2	AC125732	Rattus no
C 23	65.4	6.5	130626	2	AC113687	Rattus no
C 24	65	6.5	177801	2	AC114391	Rattus no
C 25	65	6.5	186882	2	AC095077	Rattus no
C 26	65	6.5	227009	2	AL845264	Mus muscu
C 27	64.6	6.5	208978	2	AC126593	Rattus no
C 28	64.4	6.4	125020	9	AF429315	Homo sapi
C 29	64.4	6.4	145592	2	AC119341	Homo sapi
C 30	63.8	6.4	83367	2	AC115654	Rattus no
C 31	63.8	6.4	141905	2	AC110101	Rattus no
C 32	63.6	6.4	119989	2	AC119472	Rattus no
C 33	63.6	6.4	151283	2	AC107540	Rattus no
C 34	63.4	6.3	152895	2	AC105671	Rattus no
C 35	63.2	6.3	76734	2	AC023218	Homo sapi
C 36	63.2	6.3	212355	2	AC095539	Rattus no
C 37	62.6	6.3	148481	2	AC127076	Rattus no
C 38	62.6	6.3	174605	2	AC121227	Rattus no
C 39	62.4	6.2	42748	2	AC105855	Rattus no
C 40	62.4	6.2	184402	2	AC127041	Rattus no
C 41	62.2	6.2	152732	2	AC074204	Mus muscu
C 42	62.2	6.2	182269	2	AC107416	Rattus no
C 43	62	6.2	158216	2	AC129233	Rattus no
C 44	61.8	6.2	125020	9	AF429315	Homo sapi
C 45	61.8	6.2	175440	2	AC126197	Rattus no

ALIGNMENTS

RESULT 1  
HSA297538  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
JOURNAL  
REFERENCE

HSA297538  
Homo sapiens partial RARA gene, intron 2.  
AJ297538  
AJ297538.1 GI:12054223  
Rara gene; retinoic acid receptor alpha.  
human.  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 16913)  
Cross,N.C.P. and Reiter,A.  
Unpublished  
2 (bases 1 to 16913)

linear PRI 06-JAN-2001

AUTHORS Cross,N.C.P.  
 TITLE Direct Submission  
 JOURNAL Submitted (07-AUG-2000) Cross N.C.P., Department of Haematology,  
 Imperial College School of Medicine, Hammersmith Hospital, London,  
 W12 0NN, UNITED KINGDOM

FEATURES  
 source Location/Qualifiers  
 1..16913  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="17"  
 /map="17q12"

gene 1..16913  
 /gene="RARA"  
 intron 1..16913  
 /gene="RARA"  
 /number=2

repeat\_region 1638..1947  
 /rpt\_family="AluSg"  
 repeat\_region 1949..2235  
 /rpt\_family="AluSx"  
 repeat\_region 2236..2536  
 /rpt\_family="AluSx"  
 repeat\_region 4008..4315  
 /rpt\_family="AluSx"  
 repeat\_region 4327..4619  
 /rpt\_family="AluSx"  
 repeat\_region 4669..4975  
 /rpt\_family="AluSx"  
 repeat\_region 4982..5009  
 /note="(TTTA)n"  
 repeat\_region 5011..5287  
 /rpt\_family="AluY"  
 repeat\_region 5287..5597  
 /rpt\_family="AluSx"  
 repeat\_region 5666..5776  
 /rpt\_family="MIR"  
 repeat\_region 7090..7148  
 /rpt\_family="MIR"  
 repeat\_region 7159..7326  
 /rpt\_family="AluSg/x"  
 repeat\_region 7331..7641  
 /rpt\_family="AluSg"  
 repeat\_region 8518..8659  
 /rpt\_family="L2"  
 repeat\_region 8853..9194  
 /rpt\_family="L2"  
 repeat\_region 10145..10198  
 /note="(CGG)n"

BASE COUNT 3169 a 4730 c 4922 g 4092 t  
 ORIGIN

Query Match 97.4%; Score 973.6; DB 9; Length 16913;  
 Best Local Similarity 99.0%; Pred. No. 1.5e-185;  
 Matches 990; Conservative 0; Mismatches 9; Indels 1; Gaps 1;  
 Qy 1 GTCCTGGGTAGCATGATACATTCATCCCTCCCTTTATATATGCGGGTAAATAGATAC 60  
 Db 9621 GTCCTGGGTAGCATGATACATTCATCCCTTTAGAGAGTGGGGTAAATAGATAC 9680  
 Qy 61 CCCCTCTCCAGGGTATCCCTCTTTTAGGGACCTACCCAAAGCTAGGCCCTTTCTTCCA 120  
 Db 9681 CCCCTCTCCAGGGTATCCCTCTTTTAGGGACCTACCCAAAGCTAGGCCCTTTCTTCCA 9740  
 Qy 121 GTGAACGTGATCCCGAGGGCTTCTAGATGAAGTAGTCCACTGGAGGACCAAGCTCT 180  
 Db 9741 GTGAACGTGATCCCGAGGGCTTCTAGATGAAGTAGTCCACTGGAGGACCAAGCTCT 9800  
 Qy 181 TCCCTTTATCTCCAGAGCTGGACAGTGCACAGGCGCGTACTGTCTCCAGCTAG 240  
 Db 9801 TCCCTTTATCTCCAGAGCTGGACAGTGCACAGGCGCGTACTGTCTCCAGCTAG 9860  
 Qy 241 GAGACACCTTGGGGGGCTTTGCTCGCGGGAAGCACAGAGCTGGGGAGGGGCC 300

Db 9861 GAGACACCTTGGGGGGCTTTGCTCGCGGAAGCACAGCAGCGTGGGAGGAGGCC 9920  
 Qy 301 CCTCTGCTGTGTTTGTGCAACAGCAGCAGCCGCTGCCGCTCGGGTTCGGGCGCCGGA 360  
 Db 9921 CCTCTGCTGTGTTTGTGCAACAGCAGCAGCCGCTGCCGCTCGGGTTCGGGCGCCGGA 9980  
 Qy 361 GTCACACATGATGTACAGACATGACAAAGCGGTGTCTCATTCGACACACAGCTCCG 420  
 Db 9981 GTCACACATGATGTACAGACATGACAAAGCGGTGTCTCATTCGACACACAGCTCCG 10040  
 Qy 421 AGCTGCACAATGTACACACCGGGTGCACAACTTTGGCCCCGCGCAGCCGCTTACGC 480  
 Db 10041 AGCTGCACAATGTACACACCGGGTGCACAACTTTGGCCCCGCGCAGCCGCTTACGC 10100  
 Qy 481 CTCCTGCGCGCTCTCCGGGTCTCCGGGGAGGTGGCCCGGGTTCGGCCGGGAGGGGGC 540  
 Db 10101 CTCCTGCGCGCTCTCCGGGTCTCCGGGGAGGTGGCCCGGGTTCGGCCGGGAGGGGGC 10160  
 Qy 541 TGGCGGGGAGCGCCCGGGCTGGCGAGGGGTGGCGAGGGGTGATGTACAGCGCAGCTTGCATT 600  
 Db 10161 TGGCGGGGAGCGCCCGGGCTGGCGAGGGGTGGCGAGGGGTGATGTACAGCGCAGCTTGCATT 10220  
 Qy 601 GGTCACTCGGAGGTGAGCGCCCGCAGCGAGTTTCAGCGAGAGTTTCAGCGCATTGCATT 660  
 Db 10221 GGTCACTCGGAGGTGAGCGCCCGCAGCGAGTTTCAGCGAGAGTTTCAGCGCATTGCATT 10280  
 Qy 661 AGGCAATGAGCGCGCGCTGGTGGGGTGTGTGTTAAGGGAGGAGCACCGGGACCAACC 720  
 Db 10281 AGGCAATGAGCGCGCGCTGGTGGGGTGTGTGTTAAGGGAGGAGCACCGGGACCAACC 10340  
 Qy 721 CCCCTCTTCCCGCCCGCAGCCTCTCCACACAGCTTCGCTCGCGCAGGAGTTCGCGCAGTACCA 780  
 Db 10341 CCCCTCTTCCCGCCCGCAGCCTCTCCACACAGCTTCGCTCGCGCAGGAGTTCGCGCAGTACCA 10400  
 Qy 781 AACCTTGGGGAGCGCTGGAGCGGAACTGGTATACAGGGAGGAGCGCCGCCCTCTTTC 840  
 Db 10401 AACCTTGGGGAGCGCTGGAGCGGAACTGGTATACAGGGAGGAGCGCCGCCCTCTTTC 10459  
 Qy 841 GTCCTGTCCCTCGCAGCGCCCTCTCTCCCTGACTCCCTGCTCTCTCTCTCTCTCTCTCTCTCT 900  
 Db 10460 GTCCTGTCCCTCGCAGCGCCCT 10519  
 Qy 901 TACTCTCATCTGGAGCTTTCCCT 960  
 Db 10520 TACTCTCATCTGGAGCTTTCCCT 10579  
 Qy 961 CTGCCCCCACTTGCCTGTCCACATGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1000  
 Db 10580 CTGCCCCCACTTGCCTGTCCACATGCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 10619

## RESULT 2

AC090426/c

LOCUS

DEFINITION

AC090426

AC090426

AC090426.1

HTG.

KEYWORDS

SOURCE

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 138999)

AUTHORS

TITLE

Variant-type PMU-RAR(alpha) fusion transcript in acute

promyelocytic leukemia: use of a cryptic coding sequence from

intron 2 of the RAR(alpha) gene and identification of a new

clinical subtype resistant to retinoic acid therapy

Proc. Natl. Acad. Sci. U.S.A. 99 (11), 7640-7645 (2002)

JOURNAL

MEDLINE

PUBMED

```

REFERENCE 2 (bases 1 to 138999)
AUTHORS   Gu,B., Xiong,H., Zhou,Y., Chen,B., Lu,L., Zhong,M., Yin,H.,
          Huang,W., Ren,S., Chen,S.F., Chen,Z. and Fu,G.
TITLE     Direct Submission
JOURNAL   Submitted (20-FEB-2001) Chinese National Human Genome Center at
          Shanghai, Shanghai, Shanghai 201203, P.R.China
COMMENT   -----Genome Center-----
          Center:Chinese National Human Genome Center at Shanghai
          Center code:CHGC Website: http://www.chgc.sh.cn Contact:
          fugang@chgc.sh.cn.
FEATURES  Location/Qualifiers
           source
             1..138999
               /organism="Homo sapiens"
               /db_xref="taxon:9606"
               /chromosome="17"
               /map="17q21.1"
               /clone="205ml7"
           repeat_region
             387..483
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
           repeat_region
             725..1032
               /rpt_family="SINE/Alu"
               /rpt_unit=AluJb
           repeat_region
             1033..1198
               /rpt_family="SINE/Alu"
               /rpt_unit=Alu
           repeat_region
             1217..1514
               /rpt_family="SINE/Alu"
               /rpt_unit=AluSp
           repeat_region
             complement(1534..1754)
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
           repeat_region
             2291..2366
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
           repeat_region
             complement(2607..2665)
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
           repeat_region
             3651..3905
               /rpt_family="LINE/L1"
               /rpt_unit=L1ME2
           repeat_region
             3977..4029
               /rpt_family="LINE/L1"
               /rpt_unit=L1ME2
           repeat_region
             4030..4213
               /rpt_family="LTR/MaLR"
               /rpt_unit=MSTD
           repeat_region
             4214..4509
               /rpt_family="SINE/Alu"
               /rpt_unit=AluJb
           repeat_region
             4514..4811
               /rpt_family="SINE/Alu"
               /rpt_unit=AluSx
           repeat_region
             4812..5018
               /rpt_family="LTR/MaLR"
               /rpt_unit=MSTD
           repeat_region
             5019..5102
               /rpt_family="LINE/L1"
               /rpt_unit=L1ME2
           repeat_region
             5236..5443
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
           repeat_region
             5467..5699
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
           repeat_region
             6547..6708
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
           repeat_region
             6827..6917
               /rpt_family="Low_complexity"
               /rpt_unit=GA-rich
           repeat_region
             7733..7779
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
           repeat_region
             8268..8411
               /rpt_family="Simple_repeat"
               /rpt_unit=(TGGG)n
           repeat_region
             9224..9532
               /rpt_family="SINE/Alu"
               /rpt_unit=AluY
           repeat_region
             9546..9639
               /rpt_family="Low_complexity"
               /rpt_unit=Ct-rich
           repeat_region
             10025..10426
               /rpt_family="LTR/Retroviral"
               /rpt_unit=MLT2A
           repeat_region
             10433..10471
               /rpt_family="Simple_repeat"
               /rpt_unit=(TA)n
           repeat_region
             10472..10527
               /rpt_family="LTR/Retroviral"
               /rpt_unit=MLT2A
           repeat_region
             complement(10675..10962)
               /rpt_family="SINE/Alu"
               /rpt_unit=AluY
           repeat_region
             11013..11321
               /rpt_family="SINE/Alu"
               /rpt_unit=AluY
           repeat_region
             complement(11840..12192)
               /rpt_family="LTR/MaLR"
               /rpt_unit=MSTD
           repeat_region
             12729..12835
               /rpt_family="SINE/Alu"
               /rpt_unit=FLAM_A
           repeat_region
             complement(12841..12891)
               /rpt_family="LINE/L2"
               /rpt_unit=L2
           repeat_region
             complement(14228..14548)
               /rpt_family="SINE/Alu"
               /rpt_unit=AluY
           repeat_region
             complement(14618..14701)
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
           repeat_region
             14953..15255
               /rpt_family="SINE/Alu"
               /rpt_unit=AluJo
           repeat_region
             15311..15473
               /rpt_family="Simple_repeat"
               /rpt_unit=(TA)n
           repeat_region
             15669..15981
               /rpt_family="SINE/Alu"
               /rpt_unit=AluSx
           repeat_region
             16316..16353
               /rpt_family="Simple_repeat"
               /rpt_unit=(TAA)n
           repeat_region
             complement(16863..16910)
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
           repeat_region
             17026..17324
               /rpt_family="SINE/Alu"
               /rpt_unit=AluSg
           repeat_region
             17337..17618
               /rpt_family="SINE/Alu"
               /rpt_unit=AluSx
           repeat_region
             17619..18024
               /rpt_family="LTR"
               /rpt_unit=LTR54
           repeat_region
             complement(18026..18086)
               /rpt_family="SINE/MIR"
               /rpt_unit=MIR
           repeat_region
             complement(18292..18607)
               /rpt_family="SINE/Alu"
               /rpt_unit=AluYb8
           repeat_region
             complement(18948..19264)
               /rpt_family="SINE/Alu"
               /rpt_unit=AluSx
           repeat_region
             complement(20012..20168)

```





JOURNAL

REFERENCE  
AUTHORS

Submitted (24-SEP-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 173441)  
Birken, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A., Cook, A., Cooke, P., DeArrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karatas, A., Kellis, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C., McCarthy, M., Meldrum, J., Meneus, L., Mihova, T., Mlenga, V., O'Connor, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C. H., Phunkhang, P., Poirier, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE

JOURNAL

COMMENT

Direct Submission  
Submitted (14-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 14, 2002 this sequence version replaced gi:22123315.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
<http://ftp.genome.washington.edu/RM/RepeatMasker.html>  
-----  
Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www.seq.wi.mit.edu>  
Contact: [sequence.submissions@genome.wi.mit.edu](mailto:sequence.submissions@genome.wi.mit.edu)  
-----  
Project Information  
Center project name: L11030  
Center clone name: 2267\_D\_19  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence.  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 2128: contig of 2128 bp in length  
\* 2129 2228: gap of 100 bp  
\* 2229 157931: contig of 155703 bp in length  
\* 157932 158031: gap of 100 bp  
\* 158032 173441: contig of 15410 bp in length.

FEATURES

source

1. 173441  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="CD-2267D19"

/clone\_lib="CITD1 Human BAC"  
46772 a 43344 c 40918 g 42114 t 293 others

BASE COUNT  
ORIGIN

Query Match 97.4%; Score 973.6; DB 2; Length 173441;  
Best Local Similarity 99.0%; Pred. NO. 1.2e-18;  
Matches 990; Conservative 0; Mismatches 9; Indels 1; Gaps 1;

Qy 1 GTCCTGGGTAGCATGATCCATCCCTCCCTTTATATATATGGGGTATAGATAC 60

Db 82724 GTCCCTGGGTAGCATGATCCATCCCTTTTATAGAGTGGGGTATAGATAC 82665

Qy 61 CCCCTCCCTCCAGGGTATCCCTCTTTCTAGGACCTACCCCAAGCTAGGCCCTTTCTTCA 120

Db 82664 CCCCTCCCTCCAGGGTATCCCTCTTTCTAGGACCTACCCCAAGCTAGGCCCTTTCTTCA 82605

Qy 121 GTCAACAGCTGATCCCGAGGGGCTTCTAGGATGAAGTAGTCCACTGGAAGCACACAGCTCT 180  
Db 82604 GTGAACAGTGCATCCGAGGGGCTTCTAGGATGAAGTAGTCCACTGGAAGCACACAGCTCT 82545  
Qy 181 TCCCTTTTATCTCTCCAGAGCTGGACAGTGCACACAGGGGCGGTACTGTGTTCCCCAGTAG 240  
Db 82544 TCCCTTTTATCTCTCCAGAGCTGGACAGTGCACACAGGGGCGGTACTGTGTTCCCCAGTAG 82485  
Qy 241 GAGACACCTTGGGGGGGCTTGTCTCCGGAAGACAGACAGCGTGGGGAGGAGGGCCC 300  
Db 82484 GAGACACCTTGGGGGGGCTTGTCTCCGGAAGACAGACAGCGTGGGGAGGAGGGCCC 82425  
Qy 301 CCTCTGCGCTGTGTTGTGCAACAGACACCGCGCTCCCGGTGCGGTTCGCGGGCGCGGA 360  
Db 82424 CCTCTGCGCTGTGTTGTGCAACAGACACCGCGCTCCCGGTGCGGTTCGCGGGCGCGGA 82365  
Qy 361 GTCACACATGATGTCACAGACATGACACAAGCGGTGTCTCATTCGACACAGCGTCCG 420  
Db 82364 GTCACACATGATGTCACAGACATGACACAAGCGGTGTCTCATTCGACACAGCGTCCG 82305  
Qy 421 AGCTGCACATGTCACACCCGGGTGCAACACTTGGCCCCGCGACCGGCCCTTACGC 480  
Db 82304 AGCTGCACATGTCACACCCGGGTGCAACACTTGGCCCCGCGACCGGCCCTTACGC 82245  
Qy 481 CTCCTGCCCGCTCTCCCGGTCTCCGGGGAGGTGCGCGGTTCGCGGGCAGGGGGC 540  
Db 82244 CTCCTGCCCGCTCTCCCGGTCTCCGGGGAGGTGCGCGGTTCGCGGGCAGGGGGC 82185  
Qy 541 TGGCGGGCAGGCGCGCGGGCTGGGAGCGGGTGTATGTCACGGGACAGCGTGGGGTG 600  
Db 82184 TGGCGGGCAGGCGCGCGGGCTGGGAGCGGGTGTATGTCACGGGACAGCGTGGGGTG 82125  
Qy 601 GGTACTCGAGGTGAGGCGCGCGGAGGTTCAGCGAGTTCAGCGCATTTGCATT 660  
Db 82124 GGTACTCGAGGTGAGGCGCGCGGAGGTTCAGCGAGTTCAGCGCATTTGCATT 82065  
Qy 661 AGGCAATGAGCGCGCGCTGGTGGGGTGTGTATAGGGGAGGACACCGGGACACC 720  
Db 82064 AGGCAATGAGCGCGCGCTGGTGGGGTGTGTATAGGGGAGGACACCGGGACACC 82005  
Qy 721 CCCCTTCCCGCGCGCGCGCTCCCTCCACACAGCGTTCGCTCGGCGAGGACTGACCA 780  
Db 82004 CCCCTTCCCGCGCGCGCGCTCCCTCCACACAGCGTTCGCTCGGCGAGGACTGACCA 81945  
Qy 781 AACCTTGGGGAGCTGGGAGCGGAACTGGTACAGGGGAGGACGCGCCCGCTCTTTC 840  
Db 81944 AACCTTGGGGAGCTGGGAGCGGAACTGGTACAA-GGGAGGACGCGCCCGCTCTTTC 81886  
Qy 841 GTCTTGTCTCCCTCGCAGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900  
Db 81885 GTCTTGTCTCCCTCGCAGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 81826  
Qy 901 TACTCTCATCTGGAGCTTTCCCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960  
Db 81825 TACTCTCATCTGGAGCTTTCCCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 81766  
Qy 961 CTGCCCCCACTGCTGTCACATGCCGCTCTCCCTCTC 1000  
Db 81765 CTGCCCCCACTGCTGTCACATGCCGCTCTCCCTCTC 81726

RESULT 4

AC131063/c  
LOCUS  
DEFINITION Homo sapiens chromosome 17 clone RP11-1029L16 map 17, WORKING DRAFT  
SEQUENCE, 6 unordered pieces.  
ACCESSION AC131063  
VERSION AC131063.1 GI:22267816  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.



```

QY 720 CCCCTCTTCCCGCCACACCTCTCTCCACACGGCTTCGCTCGCCGAGGACTGACC 779
Db 165606 CCCCTCTTCCCGCCACACCTCTCTCCACACGGCTTCGCTCGCCGAGGACTGACC 165547
QY 780 AAACCTTGGGGAGCTGGGAGCGGAACTGCTACAGGGAGGAGCGCCGCTCTTC 839
Db 165546 AAACCTTGGGGAGCTGGGAGCGGAACTGCTACAGGGAGGAGCGCCGCTCTTC 165488
QY 840 GGTCTTGTCCCTCGAGCCGCTCTCTCCCTGTACTCGGCGTCCCTCTGTACTCTGT 899
Db 165487 GGTCTTGTCCCTCGAGCCGCTCTCTCCCTGTACTCGGCGTCCCTCTGTACTCTGT 165428
QY 900 GTACTCTCATCTGAGCGCTTTCCTCCCTCGCTCTCTCTCTCTCTCTCTCTCCAG 959
Db 165427 GTACTCTCATCTGAGCGCTTTCCTCCCTCGCTCTCTCTCTCTCTCTCTCTCCAG 165368
QY 960 GCTGCCCCACCTGCTGTCACATGCGCGCTCTCCCTCTC 1000
Db 165367 GCTGCCCCACCTGCTGTCACATGCGCGCTCTCCCTCTC 165327

RESULT 5
AC126392/c
LOCUS
DEFINITION Homo sapiens chromosome 17 clone RP11-1029F16 map 17, WORKING DRAFT
SEQUENCE, 11 ordered pieces.
ACCESSION AC126392
VERSION AC126392.2 GI:221233095
KEYWORDS HTGS_PHASE2; HTGS_DRAFT; HTGS_FULLTOP.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 190309)
Barnes, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 190309)
Barnes, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Submitted (05-JUL-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 190309)
Barnes, N., Bastien, V., Bloom, T., Boguslavskiy, L., Boukhgalter, B.,
Camarata, J., Chang, J., Chazaro, B., Choepel, Y., Collymore, A.,
Cook, A., Cooke, P., Dearellano, K., Dewar, K., Diaz, J.S., Dodge, S.,
Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J.,
Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B.,
Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A.,
Karatas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K.,
Liu, G., MacLean, C., Macdonald, P., Major, J., Matthews, C.,
McCarthy, M., Meldrim, J., Meneus, L., Mihova, T., Mienga, V.,
Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H.,
O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K.,
Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P.,
Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P.,
Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J.,
Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H.,
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J.,
Zembek, L., Zimmer, A. and Zody, M.
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced gi:21699357.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L27553
Center clone name: 1029_F16
----- Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 187521 bases at least Q40
Consensus quality: 188193 bases at least Q30
Consensus quality: 188537 bases at least Q20
Insert size: 179000; agarose-fp
Insert size: 189309; sum-of-contigs
Quality coverage: 15.5 in Q20 bases; agarose-fp
Quality coverage: 14.6 in Q20 bases; sum-of-contigs
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 11 contigs. Gaps between the contigs
* are represented as runs of N. The order of the pieces
* is believed to be correct as given, however the sizes
* of the gaps between them are based on estimates that have
* been provided by the submitter.
* This sequence will be replaced
* by the finished sequence as soon as it is available and
* the accession number will be preserved.
* 1 30513: contig of 30513 bp in length
* 30514 30613: gap of 100 bp
* 30614 31764: contig of 1151 bp in length
* 31765 31864: gap of 100 bp
* 31865 33693: contig of 1829 bp in length
* 33694 33793: gap of 100 bp
* 33794 35405: contig of 1612 bp in length
* 35406 35505: gap of 100 bp
* 35506 37691: contig of 2186 bp in length
* 37692 37791: gap of 100 bp
* 37792 51370: contig of 13579 bp in length
* 51371 51470: gap of 100 bp
* 51471 66014: contig of 14544 bp in length
* 66015 66114: gap of 100 bp
* 66115 108183: contig of 42069 bp in length
* 108184 108283: gap of 100 bp
* 108284 130510: contig of 22227 bp in length
* 130511 130610: gap of 100 bp
* 130611 166740: contig of 36130 bp in length
* 166741 166840: gap of 100 bp
* 166841 190309: contig of 23469 bp in length.
Location/Qualifiers
1..190309
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="17"
/map="17"
/clone="RP11-1029F16"
/clone_lib="RPC1-11 Human Male BAC"
1..30513
/misc_feature
/note="assembly_fragment"

```

Phunkhang, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schuback, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

TITLE  
JOURNAL

## COMMENT

Direct Submission  
Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 6, 2002 this sequence version replaced gi:21699357.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L27553  
Center clone name: 1029\_F16  
----- Summary Statistics

Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 187521 bases at least Q40  
Consensus quality: 188193 bases at least Q30  
Consensus quality: 188537 bases at least Q20  
Insert size: 179000; agarose-fp  
Insert size: 189309; sum-of-contigs  
Quality coverage: 15.5 in Q20 bases; agarose-fp  
Quality coverage: 14.6 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 11 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.

\* 1 30513: contig of 30513 bp in length  
\* 30514 30613: gap of 100 bp  
\* 30614 31764: contig of 1151 bp in length  
\* 31765 31864: gap of 100 bp  
\* 31865 33693: contig of 1829 bp in length  
\* 33694 33793: gap of 100 bp  
\* 33794 35405: contig of 1612 bp in length  
\* 35406 35505: gap of 100 bp  
\* 35506 37691: contig of 2186 bp in length  
\* 37692 37791: gap of 100 bp  
\* 37792 51370: contig of 13579 bp in length  
\* 51371 51470: gap of 100 bp  
\* 51471 66014: contig of 14544 bp in length  
\* 66015 66114: gap of 100 bp  
\* 66115 108183: contig of 42069 bp in length  
\* 108184 108283: gap of 100 bp  
\* 108284 130510: contig of 22227 bp in length  
\* 130511 130610: gap of 100 bp  
\* 130611 166740: contig of 36130 bp in length  
\* 166741 166840: gap of 100 bp  
\* 166841 190309: contig of 23469 bp in length.

FEATURES  
source

1..190309  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="RP11-1029F16"  
/clone\_lib="RPC1-11 Human Male BAC"  
1..30513  
/misc\_feature  
/note="assembly\_fragment"



----- Summary Statistics -----  
 Sequencing vector: M13; M7815; 41% of reads  
 Sequencing vector: Plasmid; L08752; 4% of reads  
 Sequencing vector: Plasmid; n/a; 56% of reads  
 Chemistry: Dye-primer-amersham; 12% of reads  
 Chemistry: Dye-terminator Big Dye; 88% of reads  
 Assembly program: Phrap; version 0.960731  
 Consensus quality: 153948 bases at least Q40  
 Consensus quality: 156118 bases at least Q30  
 Consensus quality: 157160 bases at least Q20  
 Insert size: 63000; agarose-fp  
 Insert size: 157866; s.  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 10 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence.  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 2674: contig of 2674 bp in length  
 \* 2675 2774: gap of 100 bp  
 \* 2775 3397: contig of 623 bp in length  
 \* 3398 3497: gap of 100 bp  
 \* 3498 4542: contig of 1045 bp in length  
 \* 4543 4642: gap of 100 bp  
 \* 4643 5962: contig of 1320 bp in length  
 \* 5963 6062: gap of 100 bp  
 \* 6063 87223: contig of 81161 bp in length  
 \* 87224 87323: gap of 100 bp  
 \* 87324 95421: contig of 8098 bp in length  
 \* 95422 95521: gap of 100 bp  
 \* 95522 110722: contig of 15201 bp in length  
 \* 110723 110822: gap of 100 bp  
 \* 110823 124645: contig of 13823 bp in length  
 \* 124646 124745: gap of 100 bp  
 \* 124746 156015: contig of 31270 bp in length  
 \* 156016 156115: gap of 100 bp  
 \* 156116 158766: contig of 2651 bp in length.

## FEATURES

source

```

1. 158766
  /organism="Homo sapiens"
  /db_xref="taxon:9606"
  /chromosome="17"
  /map="17"
  /clone="RP5-1112621"
  /clone_lib="RPCI Human PAC library 5"
  1. 2674
  /note="assembly_fragment"
  clone_end:SP6
  vector_side:left
  2775..3397
  /note="assembly_fragment"
  3498..4542
  /note="assembly_fragment"
  4643..5962
  /note="assembly_fragment"
  6063..87223
  /note="assembly_fragment"
  87324..95421
  /note="assembly_fragment"
  95522..110722
  /note="assembly_fragment"
  110823..124645
  /note="assembly_fragment"
  124746..156015
  /note="assembly_fragment"
  156116..158766
  /note="assembly_fragment"
  clone_end:T7
  vector_side:right
  38051 a 38345 c 39366 g 42090 t 914 others
  ORIGIN

```

```

Query Match          95.7%; Score 956.8; DB 2; Length 158766;
Best Local Similarity 98.5%; Pred. No. 3.2e-185;
Matches 987; Conservative 0; Mismatches 12; Indels 3; Gaps 2;

QY 1  GTCCCTGGGTAGCATGTACATTTCCATCCCTCTCCTTTATATATATGGGGTAAATAGGATAC 60
    |||||
DB 72122 GTCCCTGGGTAGCATGTACATTTCCATCCCTCTCCTTTTAGAGAGTGGGGTAAATAGGATAC 72181

QY 61  CCCTCTCTCCAGGGGTATCCCTCTTTTCTAGGACCTACCCAAAGCTAGGCCTTTCTTCCA 120
    |||||
DB 72182 CCCTCTCTCCAGGGGTATCCCTCTTTTCTAGGACCTACCCAAAGCTAGGCCTTTCTTCCA 72241

QY 121 GTGAACGTGTCATCCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAGGACCAAGCTCT 180
    |||||
DB 72242 GTGAACGTGTCATCCCGAGGGCTTCTAGGATGAAGTAGTCCACTGGAGGACCAAGCTCT 72301

QY 181 TCCTTTTATCTCCAGAGCTGCACAGTGCCACAGGGGCGGTACTGTTCCCCAGCTAG 240
    |||||
DB 72302 TCCTTTTATCTCCAGAGCTGCACAGTGCCACAGGGGCGGTACTGTTCCCCAGCTAG 72361

QY 241 GAGACACCTTGGCGGGGGCTTTCTGCGCGGAAGCACAGAGCGTGGGAGGAGGGCCC 300
    |||||
DB 72362 GAGACACCTTGGCGGGGGCTTTCTGCGCGGAAGCACAGAGCGTGGGAGGAGGGCCC 72421

QY 301 CCTCTGCCCTGTGTTGTGCCAACAGCACCGCGCTGCCGCGTGGGGTTCCGGGCGGCGGA 360
    |||||
DB 72422 CCTCTGCCCTGTGTTGTGCCAACAGCACCGCGCTGCCGCGTGGGGTTCCGGGCGGCGGA 72481

QY 361 GTCACACATGATGTCACAGACATGACAAAGCGGGTGTCTCATTCGACACAGCGTCCG 420
    |||||
DB 72482 GTCACACATGATGTCACAGACATGACAAAGCGGGTGTCTCATTCGACACAGCGTCCG 72541

QY 421 AGTGTACAAATGTCACACCGGGTGTGCAAACTTTGGCCCCCGCGCACCGGCCCTACGC 480
    |||||
DB 72542 AGTGTACAAATGTCACACCGGGTGTGCAAACTTTGGCCCCCGCGCACCGGCCCTACGC 72601

QY 481 CTCCTCGCGCGCTCTCCGGGTCTCCGGGGGA--GGTGGCCGGTTCGGCGGGGCGAGGG 538
    |||||
DB 72602 CTCCTCGCGCGCTCTCCGGGTCTCCGGGGGAAGGTGGCCCGGGTTCGGCGGGGCGAGGG 72661

QY 539 GCTGGCGGGGAGCGCGCGGGGGGTGCGGAGCGGGGTATGTCAGCGGCGAGCGGGTGGG 598
    |||||
DB 72662 GCTGGCGGGGAGCGCGCGGGGGGTGCGGAGCGGGGTATGTCAGCGGCGAGCGGGTGGG 72721

QY 599 TGGGTCACTCGGAGTGAGGCGCGCGGAGGTTTCAGCGAGAGTTTCAGCCGATTTGCA 658
    |||||
DB 72722 TGGGTCACTCGGAGTGAGGCGCGCGGAGGTTTCAGCGAGAGTTTCAGCCGATTTGCA 72781

QY 659 TTAGGCAAAATGAGGCGCGCGCTGGGTGGGGGTGTGTGTTAAGGGGAGGACACCGGGACCA 718
    |||||
DB 72782 TTAGGCAAAATGAGGCGCGCGCTGGGTGGGGGTGTGTGTTAAGGGGAGGACACCGGGACCA 72841

QY 719 CCGCCCTCTTCCCGCCCGCCACACCTCTCCACACAGGGTTCGCTCGGCGAGGAGTGCAC 778
    |||||
DB 72842 CCGCCCTCTTCCCGCCCGCCACACCTCTCCACACAGGGTTCGCTCGGCGAGGAGTGCAC 72901

QY 779 CAAACCTTGGGGAGGCTGGGAGCGCGGAACTGGTACAAAGGGGAGGAGCGCGGCCCTCTT 838
    |||||
DB 72902 CAAACCTTGGGGAGGCTGGGAGCGCGGAACTGGTACAA--GGGAGGAGCGCGGCCCTCTT 72960

QY 839 CCGTCCTTGTCCCTCGAGCGCGCTCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 898
    |||||
DB 72961 CCGTCCTTGTCCCTCGAGCGCGCTCTCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 73020

QY 899 TGTACTCTCTCATCTGGAGCGCTTTCCCGCTTCTCTGCTTCTCTCTCTCTCTCTCTCTCT 958
    |||||
DB 73021 TGTACTCTCTCATCTGGAGCGCTTTCCCGCTTCTCTGCTTCTCTCTCTCTCTCTCTCTCT 73080

QY 959 GCGTCCCCCCTATGCTGCTGCACATGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1000
    |||||
DB 73081 GCGTCCCCCCTATGCTGCTGCACATGCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 73122

```

BASE COUNT 38051 a 38345 c 39366 g 42090 t 914 others  
 ORIGIN







Query Match	26.3%	Score 263.2;	DB 2;	Length 231248;
Best Local Similarity	65.1%	Pred. No. 6.4e-44;		
Matches 635;	Conservative 0;	Mismatches 248;	Indels 93;	Gaps 13;
QY 34	CTTTTATATATGGGGTAATAGGATACCC--CCTCCTCCAGGGGTATCCCTCTCTTCAGG 92			
Db 137655	CTTTTAAAGGAGAGATAATAGGATACCCCTGGTCTCTCTCGGTATCTCTTCTTCAGG 137714			
QY 93	GACCTACCCAAAGCTAGGCGCTTCTTCCAGTGAACAGTGCATCCGAGGGCTTCCTAGGATG 152			
Db 137715	GACCTACCTAACCCAGCGCTTCTTCCACTCAACAGCTGCTCTAG-----GGAGG 137765			
QY 153	AGATAGTCCACTGGAGGCAACAGCTCTTCCCTTTTATCTCCTCCAGAGCTGGACAGTGCAC 212			
Db 137766	AGTGTCTTCAACTTAAAGTC-----TCCCTTTTATCTGCTCGGGCT----- 137805			
QY 213	CAGGGCGCGGTACTGTGTTCCCACTAGGAGACACCTTGGCGGGGGCTTGTCTCGCCGGA 272			
Db 137806	----GGCCCATCTGTGTTCCCGGCTGGGAGACACCTGGCGGAGAGCTTGTCTCGCCGGA 137861			
QY 273	AGCACGCAGAGCGTGGGGAGAGGGGCCCTCTGCTCTGTGTTTGTGCCAACAGC----- 326			
Db 137862	AGCACGCAGAGTGTGGGAGAGAGCTCCCC--GCTGTGTGTTGTCTAACAGCTCAGTT 137918			
QY 327	-----ACCCGCGTGC CGCGCTCGGGTTCGGCGCCCGGAGTGCACATGATGTCA 378			
Db 137919	CTGCCGTGTGCTTATTCTCCGCGGGGTGAGCCGGCAGCTCACATGATGTCA 137978			
QY 379	GACAATGACACAAGCCGG--TGTCTCATTCGGACACAGCGTCCGAGCTGCACAATGTCA 437			
Db 137979	GACAGTGACACAAGCTTGGATGCTCTCATTCGCACACAGCGCCCAAGCTGCACAATGTCA 138038			
QY 438	CCGGGTGCCAAA-----CACTTGGCCCGCGCGACCCGGCCCTACGCCTC 483			

[illegible]

Best Local Similarity 64.8%; Pred. No. 7.7e-43;  
Matches 633; Conservative 0; Mismatches 250; Indels 94; Gaps 13;

QY 34 CTTTATATATGGGGTAATAGGATACCC-CCTCTCCAGGGATATCCCTCTTCTTAGG 92  
Db 64 CTTTTAGAGAGAGATAATAGGATACCGTCTCTCTGGATATCTCTTCTTAGG 123

QY 93 GACATACCAAGCTAGGCTTTCTTCCAGTGAACGTCATCCCGAGGCTTCTAGGATG 152  
Db 124 GACCTACTAACCAAGCTTTCTTCCACTCAACAGCTGCTCTAG-----GGAGG 174

QY 153 AAGTAGTCCACTGAAGCACCAGCTCTCTTCTTATATCTCTCAGAGCTGACAGTGCAC 212  
Db 175 AAGTGTCTCAACTTAAGTC-----TCCTTTATCTGCTCTGGCT----- 214

QY 213 CAGGGCGGCTACTGTTTCCACAGTAGGACACCTTGGCGGGCTTCTCTCGCCGA 272  
Db 215 ----GGCCCATCTCTGTTCCGGCTGGAGACACCTTGGCGGAGCTTCTCTCGCCGA 270

QY 273 AGCAGCAGAGCTGGGAGAGAGGCCCCCTCTGCTGTGTGTGTGTCACACAGC----- 326  
Db 271 AGCAGCAGAGCTGGGAGAGAGCTCCCC--GCTGTGTGTGTGCTTAACAGCTCAGTT 327

QY 327 -----ACCCGCGCTGCGCGCTGGGTTCCGGGGCGGAGTCACACATGATGTACAC 378  
Db 328 CTGCGGTGCTCTTATTCCCGCGGCTGAGCCGCGAGCTCCACATGATGTACAC 387

QY 379 GACATACACACAGCCGG-TGTCATCTCCAGACACAGGTCGAGCTGCACAAATGTACAC 437  
Db 388 GACAGTACACAGCTGGATGTCTCATTCGACACAGCGCCCAAGCTGCACAAATGTACAC 447

QY 438 CCGGGTGCCAAA-----CACTTGGCCCGCGCGAGCCCGGCGCTACGCGTC 483  
Db 448 CCGGGACACCCAGCATGCTGCTGCTGCTTAAAGCTGGGCGACCTGCAATTGGGCC 507

QY 484 CTGCGCGCGCTCTCGCGCTCTCGGGGAGAGTGCCCGGTTTCGCGCGGCGAGGGGCTGG 543  
Db 508 CTGTGCTACTGCGCGCTGCTTTCGAGATTGGGGCGCGACAGGTCGAGATGGGGCCAG 567

QY 544 CGGGCAGCCCGCGCGCTGCGGAGCGGCTGATGTACAGGCGACGCTGGTGGGT 603  
Db 568 TGGCGAGCCCGCGAGCGGGTTGGCGAGCTGGTGTGTACAGGCGAGCGGTGGGTGGGT 627

QY 604 CACTCGAGGTAGGCGCCGCCAGCGAGTTCAGCGAGAGTTCAGCGCATTTGCATTAGG 663  
Db 628 CAGTCGAGGTAGGCGCCGCCAGCGAGTTCACAGAGATTTCAGCGCAATTGCATTAGG 687

QY 664 CAATAGGCGCGGCC--TGCGTGGGGTGTGTTTAAAGGGAGGACACCGGACACCC 721  
Db 688 CAATAGAGCCGGGCTTTGGGGGGGGGGTGTGTTAGGGGGAGGACACCGGACACCC 747

QY 722 CCCTCTCCCGCCCGCCACCTCTCCACAGGCTTCCGCTCGCGGAGGACTGACCA- 780  
Db 748 C-----ACCTCCCGCGCTGCTATCTTCTGCGAGGGGATTGACCA 788

QY 781 AACCTTTGGGAGCTGGGAGCGGAACTGGTACAAGGGAGGAGCGCGCGCTCTCTCC 840  
Db 789 AAGCACTAGGAGCTGGGAGCGGAACTGG--GATGGGAGAGATGCCCTACTCTTCG 846

QY 841 GTCTTGTCCCTCGCAGCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 900  
Db 847 GTCC----AGCCCTCCAGTGCACCCCTTATTCATATCCAGCCAGGTTCATGACACAGC 902

QY 901 TACTCTCTATCTGAGGCTTTCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 960  
Db 903 CATTTCCCGCTCTGTCT 962

QY 961 CTGCCCCCACTTGGCTG 977  
Db 963 TGGCCCTCGCTTCCCTG 979

RESULT 10

AX344791/c AX344791 13508 bp DNA linear PAT 01-FEB-2002  
LOCUS Sequence 216 from Patent WO0200927.  
DEFINITION AX344791  
ACCESSION AX344791  
VERSION AX344791.1 GI:18492677  
KEYWORDS synthetic construct.  
SOURCE synthetic construct.  
ORGANISM artificial sequences.  
REFERENCE 1  
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.  
TITLE Diagnosis of diseases associated with development genes  
JOURNAL Patent: WO 0200927-A 216 03-JAN-2002;  
EpiGenomics AG (DE)  
FEATURES Location/Qualifiers  
source 1..13508  
/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 3338 a 286 c 3789 g 6094 t 1 others  
ORIGIN  
Query Match 24.3%; Score 242.8; DB 6; Length 13508;  
Best Local Similarity 80.3%; Pred. No. 1.2e-39;  
Matches 297; Conservative 0; Mismatches 72; Indels 1; Gaps 1;  
QY 631 AGTTCAGCGAGATTGACGGCGGATTCGATTAGGCAATAGCGCGGCTGGTGGGGT 690  
Db 665 AATTCACGAAATTCACCGCATTTACATTAACAATAAACCACCGACCTAAATAAAAT 6606  
QY 691 GTGTCTTAAGGGAGACACGGGACACCGCCCTCTTCCCGCCCGCCACCTCTCTCT 750  
Db 6605 ATATATTAATAAATAAACAACCGAAACCGCCCTCTTCCCGCCCGCCACCTCTCTCT 6546  
QY 751 CCACGGCTTCGCTCGCGCAGGAGTACCAAACTTGGGGGAGCTTGGAGCGCGGAATG 810  
Db 6545 CCACGACTTCGCTCGACCAAACTTAACCAAACTTAACCAAACTTAACCAAACTTA 6486  
QY 811 GTACAGGGGAGGAGCGCGCGCTTCTTCCGCTTCTTCCGCTTCTTCCGCTTCTCT 870  
Db 6485 ATACAA-AAAAAAGCGCGCGCGCTTCTTCCGCTTCTTCCGCTTCTTCCGCTTCTCT 6427  
QY 871 CCTGTACTCGCGCT 930  
Db 6426 CCTATCTCGAGTCCCT 6367  
QY 931 TGCTTCT 990  
Db 6366 TACTTCT 6307  
QY 991 TCTCCCTCTC 1000  
Db 6306 TCTCCCTCTC 6297

RESULT 11  
AK098172  
LOCUS Homo sapiens cDNA FLJ40853 fis, clone TRACH2015486, highly similar  
DEFINITION to RETINOIC ACID RECEPTOR ALPHA.  
ACCESSION AK098172  
VERSION AK098172.1 GI:21758129  
KEYWORDS oligo capping; fis (full insert sequence).  
SOURCE Homo sapiens trachea cDNA to mRNA, clone lib:TRACH2  
clone:TRACH2015486.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1  
AUTHORS Suzuki, O., Sasaki, N., Aotsuka, S., Shoji, T., Ichihara, T.,  
Shiohata, N., Matsumoto, K., Hirano, M., Sano, S., Nomura, R.,  
Yoshikawa, Y., Matsumura, Y., Moriya, S., Chiba, E., Momiyama, H.,  
Onogawa, S., Kaeriyama, S., Satoh, N., Matsunawa, H., Takahashi, E.,

Kataoka, R., Kuga, N., Kuroda, A., Satoh, I., Kamata, K., Takami, S., Terashima, Y., Watanabe, M., Sugiyama, T., Irie, R., Otsuki, T., Sato, H., Ota, T., Wakamatsu, A., Ishii, S., Yamamoto, J., Isono, Y., Kawai-Hio, Y., Saito, K., Nishikawa, T., Kimura, K., Yamashita, H., Matsuo, K., Nakamura, Y., Sekine, M., Kikuchi, H., Kanda, K., Watsushima, M., Murakawa, K., Kanehori, K., Takahashi-Fujii, A., Oshima, A., Sugiyama, A., Kawakami, B., Suzuki, Y., Sugano, S., Nagahari, K., Masuho, Y., Nagai, K. and Isogai, T.  
NEDO human cDNA sequencing project

**TITLE**  
Unpublished  
**JOURNAL**  
2 (bases 1 to 2008)  
**REFERENCE**  
Isogai, T. and Yamamoto, J.  
**AUTHORS**  
Direct Submission  
**TITLE**  
Submitted (04-JUL-2002) Takao Isogai, FLJ Project (HRI Team); 2-6-7  
**JOURNAL**  
kazusa-kamata, Kisarazu, Chiba 292-0812, Japan  
(E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
**COMMENT**  
NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology (RAB); cDNA library construction: Helix Research Institute (HRI) (supported by Japan Key Technology Center etc.); 5'- & 3'-end one pass sequencing: RAB, HRI, and Biotechnology Center, National Institute of Technology and Evaluation; clone selection for full insert sequencing: HRI and RAB; annotation: HRI and RAB.

**FEATURES**  
Location/Qualifiers  
source  
1. .2008  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="TRACH2015486"  
/tissue\_type="trachea"  
/clone\_lib="TRACH2"  
/note="cloning vector: pME18SFL3"  
BASE COUNT 405 a 669 c 576 g 358 t  
ORIGIN

Query Match 22.8%; Score 228; DB 9: Length 2008;  
Best Local Similarity 100.0%; Pred. No. 1.5e-36;  
Matches 228; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 366 ACATGATGTCACAGCAATGACAAAGCGGTGTCATTCGACACAGCGCCGAGCTG 425  
|||||  
Db 1 ACATGATGTCACAGCAATGACAAAGCGGTGTCATTCGACACAGCGCCGAGCTG 60  
|||||

QY 426 CACAATGTCACCGGGTGCCAAACACTTGGCCCCGCGACCGCCCTACGCTCTCT 485  
|||||  
Db 61 CACAATGTCACCGGGTGCCAAACACTTGGCCCCGCGACCGCCCTACGCTCTCT 120  
|||||

QY 486 GCCCGCTCTCCGCTCCGGGGAGGTGGCCGGTTCGGCCGGCAGGGGCTGGCG 545  
|||||  
Db 121 GCCCGCTCTCCGCTCCGGGGAGGTGGCCGGTTCGGCCGGCAGGGGCTGGCG 180  
|||||

QY 546 GCGGAGCCCCGCGGCTGGGAGCGGTGATGTCACGGGACGG 593  
|||||  
Db 181 GCGGAGCCCCGCGGCTGGGAGCGGTGATGTCACGGGACGG 228  
|||||

**RESULT 12**  
AX344790  
**LOCUS** AX344790 13508 bp DNA linear PAT 01-FEB-2002  
**DEFINITION** Sequence 215 from Patent WO0200927.  
**ACCESSION** AX344790  
**VERSION** AX344790.1 GI:18492676  
**KEYWORDS**  
synthetic construct.  
**SOURCE** synthetic construct  
**ORGANISM** artificial sequences.  
**REFERENCE** 1  
**AUTHORS** Olek, A., Piepenbrock, C. and Berlin, K.  
**TITLE** Diagnosis of diseases associated with development genes  
**JOURNAL** Patent: WO 0200927-A 215 03-JAN-2002;  
**Epigenomics** AG (DE)  
**FEATURES** Location/Qualifiers  
source  
1. .19508

/organism="synthetic construct"  
/db\_xref="taxon:32630"  
/note="chemically treated genomic DNA (Homo sapiens)"  
BASE COUNT 2350 a 286 c 4030 g 6841 t  
ORIGIN

Query Match 14.4%; Score 143.8; DB 6: Length 13508;  
Best Local Similarity 63.2%; Pred. No. 1.8e-19;  
Matches 237; Conservative 0; Mismatches 137; Indels 1; Gaps 1;

QY 626 AGCGAGTTTCAGGAGTTCAGCGCATTCATTAGGCAATGAGCCCGCTGGGTG 685  
|||||  
Db 6839 AGGGTAGTTACGAGAGTTAGTCGATGATTAGGTAAATGAGTTGGGTTGGGTG 6898  
|||||

QY 686 GGGGTGTGTAAAGGGAGGACCGGGACACCCCTCTTCCCGCCGCCACACCTC 745  
|||||  
Db 6899 GGGGTGTGTAAAGGGAGGATATCGGGATTATTTTTTTTTTTTCGTTTATTATT 6958  
|||||

QY 746 CTCACACCGGCTTCGCTCGCGGAGGACTGACCAACCTTGGGGAGCTGGAGCGCG 805  
|||||  
Db 6959 TTTTATTACGGTTTCGTTTCGGTTAGGATGATTAAATTTGGGGAGTTGGGATCGG 7018  
|||||

QY 806 AACTGGTACAAGGGAGGACGCGCCCTCTTCCCTCTTCCCTCGCAGCCCTC 865  
|||||  
Db 7019 AATTGGTATAA-GGAGGAGCTTCGTTTTTTTCGTTTTCGTTTTCGTTTTCG 7077  
|||||

QY 866 CTCCTCCCTGTTACTCGCGCTCTCTGTACTCTGTGTACTCTGTGTACTCTGTGT 925  
|||||  
Db 7078 TTTTTCCTGTTACTCGCGCTCTCTGTACTCTGTGTACTCTGTGTACTCTGTGT 7137  
|||||

QY 926 CTCCTGCT 985  
|||||  
Db 7138 TTTTTCCTGTTACTCGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 7197  
|||||

QY 986 CGCGCTCTCCCTCTC 1000  
|||||  
Db 7198 TCGTTTTTTTTTTTC 7212  
|||||

**RESULT 13**  
AC111746/c  
**LOCUS** AC111746 162393 bp DNA linear HTG 13-JUL-2002  
**DEFINITION** Rattus norvegicus clone CH230-15011, \*\*\* SEQUENCING IN PROGRESS  
\*\*\*, 72 unordered pieces.  
**ACCESSION** AC111746  
**VERSION** AC111746.2 GI:21736893  
**KEYWORDS** HTG; HTGS, PHASE1.  
**SOURCE** Norway rat.  
**ORGANISM** Rattus norvegicus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.  
1 (bases 1 to 162393)  
Muzny, D.M., Adams, C., Adio-Oduola, B., Ali-osman, F.R., Allen, C.,  
Alsbrooks, S.L., Amaratunge, H.C., Are, J.R., Ayale, M., Banks, T.,  
Barbaria, J., Benton, J., Bimage, K., Blankenburg, K., Bonnin, D.,  
Bouck, J., Bowie, S., Brivea, M., Brown, E., Brown, N.P., Bryant, N.P.,  
Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
Carron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
Chen, G., Chen, R., Chen, Z., Chowdhry, I., Christopheoulos, C.,  
Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
Davilla, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
Douthwaite, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
Earnhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,  
Gabisi, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
Gorrell, J.H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
Harris, C., Harris, K., Hart, M., Havlak, P., Hawes, A., Hernandez, J.,  
Hernandez, O., Hodgson, A., Hogue, M., Holloway, C., Hollins, B.,  
Honsi, F., Howard, S., Huber, J., Hulyk, S., Hume, J., Jackson, L.E.,  
Jacobson, B., Jia, Y., Johnson, R., Jollivet, S., Joudah, S.,  
Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,

Kratovic, J., Kureshi, A., Landry, N., Leal, B., Lewis, L.C., Lewis, L., Li, J., Li, Z., Licharge, O., Lieu, C., Liu, J., Liu, W., Lounseged, H., Lozano, R.J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J., Maheshwari, M., Mapua, P., Martin, R., Martindale, A., Martinez, E., Massey, E., Hawthiney, E., McLeod, M.P., Meador, M., Mei, G., Metzker, M., Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S., Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N., Nguyen, N., Nickerson, E., Nwokenkwo, S., Ogih, M., Okwuonu, G., Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L., Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savery, G., Scherger, S., Scott, G., Shen, H., Shoohtari, N., Sisson, I., Sutton, A., Svatka, T., Sparks, A., Stanley, H., Stone, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R., Wang, Q., Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D., Weinstein, G., and Gibbs, R.

## TITLE

JOURNAL

## REFERENCE

2 (bases 1 to 162393)

## AUTHORS

TITLE

JOURNAL

Submitted (19-FEB-2002)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (13-JUL-2002)

Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

## COMMENT

On Jul 12, 2002 this sequence version replaced gi:18701588.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)

----- Project Information

Center project name: GOAP

Center clone name: CH230-15011

----- Summary Statistics

Sequencing vector: Plasmid;

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990329

Consensus quality: 96609 bases at least Q40

Consensus quality: 101360 bases at least Q30

Consensus quality: 105362 bases at least Q20

-----

\* NOTE: Estimated insert size may differ from sequence length

\* (see [http://www.hgsc.bcm.tmc.edu/docs/Genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html)).

\* NOTE: This is a 'working draft' sequence. It currently

\* consists of 72 contigs. The true order of the pieces

\* is not known and their order in this sequence record is

\* arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence

\* as soon as it is available and the accession number will

\* be preserved.

\* 1 1242: contig of 1242 bp in length

\* 1243 1342: gap of unknown length

\* 1343 2427: contig of 1085 bp in length

\* 2428 2527: gap of unknown length

\* 2528 3635: contig of 1108 bp in length

\* 3636 3735: gap of unknown length

\* 3736 4821: contig of 1086 bp in length

\* 4822 4921: gap of unknown length

\* 4922 6640: contig of 1719 bp in length

\* 6641 6740: gap of unknown length

\* 6741 8216: contig of 1476 bp in length

\* 8217 8316: gap of unknown length

8317 9346: contig of 1030 bp in length

9347 9446: gap of unknown length

9447 10691: contig of 1245 bp in length

10692 10791: gap of unknown length

10792 11851: contig of 1060 bp in length

11852 11951: gap of unknown length

11952 13509: contig of 1558 bp in length

13510 13609: gap of unknown length

13610 14869: contig of 1260 bp in length

14870 14969: gap of unknown length

14970 16275: contig of 1306 bp in length

16276 16375: gap of unknown length

16376 17440: contig of 1065 bp in length

17441 17540: gap of unknown length

17541 18956: contig of 1416 bp in length

18957 19056: gap of unknown length

19057 20611: contig of 1005 bp in length

20612 20662 20161: gap of unknown length

20663 21232: contig of 1071 bp in length

21233 21332: gap of unknown length

21333 22378: contig of 1046 bp in length

22379 22478: gap of unknown length

22479 23629: contig of 1151 bp in length

23630 23729: gap of unknown length

23730 24993: contig of 1264 bp in length

24994 25094 25648: contig of 1555 bp in length

25095 26649: gap of unknown length

26649 28383: contig of 1635 bp in length

28384 28483: gap of unknown length

28484 29659 29758: gap of unknown length

29759 31549: contig of 1790 bp in length

31549 32915: contig of 1267 bp in length

32916 33016 33015: gap of unknown length

33016 34132: contig of 1117 bp in length

34133 34233 35376: contig of 1144 bp in length

34233 35377 35476: gap of unknown length

35377 35477 36835: contig of 1359 bp in length

35477 36836 36935: gap of unknown length

36936 38119: contig of 1184 bp in length

38120 38219: gap of unknown length

38220 40148: contig of 1829 bp in length

40149 41648: gap of unknown length

41649 41748: contig of 1500 bp in length

41749 44023: gap of unknown length

44024 44123: contig of 2275 bp in length

44124 45297: gap of unknown length

45297 45398 45397: contig of 1174 bp in length

45398 46533: contig of 1136 bp in length

46534 46633: gap of unknown length

46634 48451: contig of 1818 bp in length

48452 48551: gap of unknown length

48552 49889: contig of 1338 bp in length

49890 49989: gap of unknown length

49990 52066: contig of 2077 bp in length

52067 52166: gap of unknown length

52167 53724: contig of 1558 bp in length

53725 53824: gap of unknown length

53825 55350: contig of 1526 bp in length

55351 57835: gap of unknown length

57836 57935: gap of unknown length

57936 59470: contig of 1535 bp in length

59471 59570: gap of unknown length

59571 61049: contig of 1479 bp in length

61050 61149: gap of unknown length

61150 62619: contig of 1470 bp in length

62620 65125: contig of 2406 bp in length



```

Query Match      7.3%; Score 72.6; DB 6; Length 7218;
Best Local Similarity 11.5%; Pred. NO. 6.2e-05;
Matches 27; Conservative 142; Mismatches 66; Indels 0; Gaps 0;

QY 766 GCCAGGACTGACCAACCTTGGGGAGCCTGGGAGCCGGAACCTGGTACAGGGGAGGAC 825
    |||| || || || |||| || || || || || || || || || || || || ||
Db 1010 GCCATACGCTCACAGAAATAATCCGAGCTTGGCTGCAGGTCGAGGAGCTTGCATYYY 1069

QY 826 GCCCGCCCTCTTCGTCTCTGTCCTCCCTCGCAGCCCTCCTCTCTCCTCTCTCTCTCTC 885
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1070 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1129

QY 886 CCTGTGACTGTGTACTCTCATCTGAGCCTTCCCTCTCTCTCTCTCTCTCTCTCTCTC 945
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1130 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1189

QY 946 CTCCCTCTCCAGGCTGCCCTCTGTCACATGCCGCTCTCCCTCTC 1000
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1190 YYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYYY 1244

```

Search completed: March 30, 2003, 12:31:19  
 Job time : 2702.41 secs

GenCore version 5.1.4.p5.4578  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: March 30, 2003, 07:57:38 ; Search time 196.103 Seconds  
(without alignments)

11483.757 Million cell updates/sec

Title: US-09-691-220-3\_COPY\_1\_1000

Perfect score: 1000

Sequence: 1 gtccttggttagcatgtaca.....acatgcgccttcctcttc 1000

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 1125999159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

ID	Score	Query Match	Length	DB ID	Description
1:	1000	100.0	20512	24	AA080208
2:	242.8	24.3	13508	24	AA0202789
3:	143.8	14.4	13508	24	AA070516
4:	59.2	5.9	53522	24	AB052496
5:	59.2	5.9	53526	19	AA052497
6:	59.2	5.9	53577	17	AA025519
7:	59.2	5.9	53577	17	AA025519
8:	56.8	5.7	114955	20	AA025519
9:	54.4	5.4	114955	20	AA025519

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1000	100.0	20512	24	AA080208
2	242.8	24.3	13508	24	AA0202789
3	143.8	14.4	13508	24	AA070516
4	59.2	5.9	53522	24	AB052496
5	59.2	5.9	53526	19	AA052497
6	59.2	5.9	53577	17	AA025519
7	59.2	5.9	53577	17	AA025519
8	56.8	5.7	114955	20	AA025519
9	54.4	5.4	114955	20	AA025519

#### ALIGNMENTS

##### RESULT 1

AA080208  
ID AAL38339 standard; DNA; 20512 BP.

XX AC AAL38339;

XX AC AAL38339;

DT 15-AUG-2002 (first entry)

XX Genomic DNA encoding the human nuclear hormone receptor protein.  
DE DE Human nuclear hormone receptor; therapeutic agent; gene therapy;  
XX KW Immune response; chromosome 17; single nucleotide polymorphism;  
KW gene; ds.

XX Homo sapiens.

XX Key Location/Qualifiers

FT CDS 1684...15208

FT FT /tag= a

FT FT /product= "Human nuclear hormone receptor protein"

FT FT 1684...1846

FT FT /tag= b

FT FT /number= 1

FT FT 1847...6808

FT FT /tag= c

FT FT /number= 1

FT FT 6809...6957

FT FT /tag= d

FT FT /number= 2

FT FT 6958...8277

FT FT /tag= e

FT FT /number= 2

Human ovarian and  
Human reproductive  
Human reproductive  
Pseudorabies virus  
Oligonucleotide fo  
Oligonucleotide fo  
Human ovarian tumo  
Human herpesvirus  
Human gene express  
Primer used in the  
Human cDNA sequenc  
Primer used in the  
Human colon cancer  
Human colon cancer  
Adenyl cyclase t  
Listeria monocytog  
Human IL-1ra BAC c  
Gene #3953 used to  
Lung cancer relate  
Lung cancer relate  
Oligonucleotide fo  
Oligonucleotide fo  
DNA encoding endot  
Primer used in the  
Primer used in the  
Listeria monocytog  
Listeria monocytog  
Listeria monocytog  
Oligonucleotide fo  
Oligonucleotide fo  
Oligonucleotide fo  
Hsv-1 IR-L (positi  
Rabbit low density  
Human ovarian and





Db 601 GGTCACTCGAGGTGAGCGCGCCAGGCGAGTTTCAGCGAGAGTTTCAGCGCATTCATT 660  
 Qy 661 AGGCAATGAGCGCGCGCTGGTGGGGTGTGTAGGGAGGAGACACCGGGACCC 720  
 Db 661 AGGCAATGAGCGCGCGCTGGTGGGGTGTGTAGGGAGGAGACACCGGGACCC 720  
 Qy 721 CCCCTCTCCCGCCCGCCACCTCTCCACCGGCTTCGCTCGCGGAGGAGTACCA 780  
 Db 721 CCCCTCTCCCGCCCGCCACCTCTCCACCGGCTTCGCTCGCGGAGGAGTACCA 780  
 Qy 781 AACCTTGGGAGCTGGAGCGGCACTGTACAGGAGGAGCGCGCCCTCTTCC 840  
 Db 781 AACCTTGGGAGCTGGAGCGGCACTGTACAGGAGGAGCGCGCCCTCTTCC 840  
 Qy 841 GTCCTGTGCTCGACGCGCCCTCTCTCGCTGACTCGGGTCCCTGTACTCTGTG 900  
 Db 841 GTCCTGTGCTCGACGCGCCCTCTCTCGCTGACTCGGGTCCCTGTACTCTGTG 900  
 Qy 901 TACTCTCATCTGGAGCCTTTCCCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 960  
 Db 901 TACTCTCATCTGGAGCCTTTCCCGCTTCTCTCTCTCTCTCTCTCTCTCTCTCT 960  
 Qy 961 CTGCCCCACTTGCTGTCCACATGCGCGCTCTCCCTCTC 1000  
 Db 961 CTGCCCCACTTGCTGTCCACATGCGCGCTCTCCCTCTC 1000

RESULT 2

ABN80199/c  
 ID ABN80199 standard; DNA; 13508 BP.  
 AC  
 AC ABN80199;  
 XX  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Human chemically modified disease associated gene SEQ ID NO 216.  
 XX  
 KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
 KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
 KW antidiabetic; cytostatic; anticonvulsant; ds.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO200200927-A2.  
 XX  
 .PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07536.  
 XX  
 PR 30-JUN-2000; 2000DE-1032529.  
 PR 01-SEP-2000; 2000DE-1043826.  
 XX  
 PA (EPIG-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-130908/17.  
 XX  
 XX Novel nucleic acid useful for diagnosis and therapy of diseases  
 PT associated with development genes such as diabetes, comprises a  
 PT sequence of a segment of chemically pretreated DNA of genes associated  
 PT with development  
 XX  
 PS Claim 1; SEQ ID NO 216; 27pp; English.  
 XX  
 CC The invention relates to a nucleic acid (I) comprising a sequence at  
 CC least 18 bases in length of a segment of chemically pretreated DNA (II)  
 CC of genes associated with development selected from 87 genes listed in  
 CC the specification such as ACCPN, ADFN, or AFDI and comprising one of 350  
 CC sequences (ABN79984-ABN80333) or their complements. The invention is  
 CC useful for the diagnosis or therapy of diseases associated with

CC development genes, in particular disease related to homeobox containing  
 CC genes (HOX), like diabetes, cancer, apoptosis related diseases, syndromes  
 CC associated with congenital heart disease, epilepsy, diseases related to  
 CC histone deacetylation, Currarino syndrome, diseases related with the  
 CC development of the brain and limb girdle muscular dystrophy and dwarfism.  
 CC Oligomers specific to each of the genes are useful for detecting the  
 CC methylation state of all CpG dinucleotides within the 350 sequences or  
 CC (II) and their complementary sequences, as primer oligonucleotides for  
 CC the amplification of the 350 sequences, (iii) and/or their complements and  
 CC as oligomer probes for detecting the cytosine methylation state and/or  
 CC single nucleotide polymorphisms (SNPs).  
 CC Note: The sequence data for this patent did not form part of the printed  
 CC specification but is based on sequence information supplied to Derwent by  
 CC the European Patent Office.

SQ Sequence 13508 BP; 3338 A; 286 C; 3789 G; 6094 T; 1 other;

Query Match 24.3%; Score 242.8; DB 24; Length 13508;  
 Best Local Similarity 80.3%; Pred No. 8.3e-45;  
 Matches 297; Conservative 0; Mismatches 72; Indels 1; Gaps 1;  
 QY 631 AGTTGAGGAGAGTTCAGCGCATTCATTAGGCAATAGAGCGCGCTGGTGGGGT 690  
 Db 6665 AATTCACGAAATTCACCGCATTCATTAAACAATAAAACCGACCTAAATAAAAT 6606  
 QY 691 GTGTGTAAAGGGAGAGACACCGGAGCCACCCCTCTTCCCGCCGACCACTCTCCA 750  
 Db 6605 ATATATTTAAAAAAAACACCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 6546  
 QY 751 CCACGGCTTCGCTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 810  
 Db 6545 CCACGACTTCGCTCGCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 6486  
 QY 811 GTACAGGGGAGGAGCGCGCCCTCTTCGCTCTGTGTCGCGCGCGCGCGCGCTCTC 870  
 Db 6485 ATACAA-AAAAAAGCGCGCGCGCGCGCTTCGCTCTATCCCTTCGCAACCCCTCTC 6427  
 QY 871 CCGTACTCGGCGTCCCTCTGTACTCTGTGTACTCTCTCATCTGAGCCCTTCCCTCTCC 930  
 Db 6426 CCGTACTCGGCGTCCCTCTGTACTCTGTGTACTCTCTCATCTGAGCCCTTCCCTCTCC 6367  
 QY 931 TCGTCT 990  
 Db 6366 TACTTCT 6307  
 QY 991 TCT 1000  
 Db 6306 TCT 6297

RESULT 3

ABN80198  
 ID ABN80198 standard; DNA; 13508 BP.  
 XX  
 AC ABN80198;  
 XX  
 XX  
 DT 15-JUL-2002 (first entry)  
 XX  
 DE Human chemically modified disease associated gene SEQ ID NO 215.  
 XX  
 KW Human; development; homeobox gene; HOX; diabetes; cancer; apoptosis;  
 KW heart disease; epilepsy; histone deacetylation; muscular dystrophy;  
 KW dwarfism; single nucleotide polymorphism; SNP; cytosine methylation;  
 KW antidiabetic; cytostatic; anticonvulsant; ds.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO200200927-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP07536.

[illegible]

RESULT 4  
AAD30228  
ID AAD30228 standard; DNA; 53522 BP.  
XX AC  
XX AAD30228;  
XX DT 17-MAY-2002 (first entry)  
XX XX Human PKD1 gene.  
XX DE  
XX KW Human; PKD1 gene; autosomal dominant polycystic kidney disease; ADPKD;  
XX KM acquired cystic disease; transgenic animal; chromosome 16; ds.  
XX OS Homo sapiens.  
XX PN WO200206529-A2.  
XX PD 24-JAN-2002.  
XX PF 13-JUL-2001; 2001WO-US22035.  
XX PR 13-JUL-2000; 2000US-218261P.  
XX PR 13-APR-2001; 2001US-283691P.  
XX PA (UWJO ) UNIV JOHNS HOPKINS SCHOOL MEDICINE.  
XX PI Germino GG, Watnick TJ, Phakdeekitcharoen B;  
XX DR WPI; 2002-179805/23.  
XX PT Novel primer for diagnosing polycystic kidney disease-associated  
PT disorder, comprises regions having sequence that selectively hybridizes  
PT to polycystic kidney disease gene sequence -  
XX Claim 20; Page 127-156; 192pp; English.  
XX The present invention relates to compositions and methods useful for the  
CC identification and detection of polycystic kidney disease (PKD1) gene  
CC mutations. The invention also relates to primers comprising a 5' region  
CC having a sequence that selectively hybridises to a PKD1 gene sequence  
CC and optionally, to a PKD1 homologue sequence and an adjacent 3' region  
CC having a sequence that selectively hybridises to a PKD1 gene sequence  
CC and not to a PKD1 homologue sequence. Primer pairs of the invention are  
CC useful for detecting the presence or absence of a mutation in a PKD1  
CC polynucleotide in a sample, for identifying a subject at risk for a  
CC PKD1-associated disorder such as autosomal dominant polycystic kidney  
CC disease (ADPKD) or acquired cystic disease and for diagnosing a PKD1-  
CC associated disorder in a subject. They are useful for selectively  
CC amplifying a region of a PKD1 gene. PKD1 DNA fragments are useful  
CC detecting the presence of a mutant PKD1 polynucleotide in a sample,  
CC as a probe for an amplification reaction, in hybridisation or  
CC amplification assays of biological samples to detect abnormalities  
CC of PKD1 expression and for engineering transgenic animals. The present  
CC sequence is human PKD1 gene located on chromosome 16.  
XX SQ Sequence 53522 BP; 8486 A; 17664 C; 15766 G; 11606 T; 0 other;  
  
Query Match 5.9%; Score 59.2; DB 24; Length 53522;  
Best Local Similarity 58.5%; Pred. No. 0.0011;  
Matches 103; Conservative 0; Mismatches 73; Indels 0; Gaps 0;  
  
QY 825 CGCGGCGCCTCTTCGGTCTGTGCCCTCGAGCCCCCTCTCTCCCTGTACTCGGCGT 884  
Db 35409 CCCCCTCTCCCATCCCCCTCTCTCTCCCACTCCTCTCTGCCACATCCCTCTCTC 35468  
  
QY 885 CCCTCTGTACTGTGTACTCCCTCATCTGGAGCGTTTCCCGCTTCTTGCTCTCTCTCT 944  
Db 35469 CCTCATCCCCCTCTCTCTCTCCCTCCCTCTCTCTCTCTCTCTCTCTCTCTCTCT 35528  
  
QY 945 CTCTCGCCCTTCCAGGCTGCCGCCACTTGCCTGTGCCACATCGCGCTCTCTCTCTCT 1000  
Db 35529 CATTCCCGCTCTTCCCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 35584





QY 190 CTCCTCCAGAGTGGACAGTGCACAGGCGCGGTACTGTTCCCTCCAGCTAGGAGACACT 249  
Db 104858 HNNNSCGGCGCGGNNHNNNSCCVGGCCVCGGNNHNNNSCCVGGCCVCGGNNHNN 104799  
QY 250 TGGCGGGGCTTGTCTCCGCGAGAGCAGAGCGTGGGAGGAGGCGCCCTCTGCT 309  
Db 104798 NSGCCVGGCGGNNHNNNSCGCCVGGCGGNNHNNNSCGCCVGGCGGCG 104739  
QY 310 GTGTTGTGCAACAGCACCCGCTGCGCGGTGCTGCGGCGCGGAGTCAACAT 369  
Db 104738 GNNHNNNSCGCGCGGNNHNNNSCGCGCGGCGGCGGNNHNNNSCGG 104679  
QY 370 GATGTCACAGCAATGACACAGCGGTGCTCTATTCGACACAGCGTCCAGAGTCACA 429  
Db 104678 GCGCGCGGCGGNNHNNNSCGCGCGGCGGCGGNNHNNNSCGCGCGGCG 104619  
QY 430 ATGTCACACCGGCTGCAACACTTGGCGCGCGGCGGCGGCGGCGGCGGCGG 489  
Db 104618 CVGGC-CVGGGNNHNNNSGCGCGCGGCGGCGGCGGCGGNNHNNNSCGGCGGCGG 104560  
QY 490 CCGCTCTCCGCTCCGCGGAGGTGCGCGGTTCGCGCGGCGGAGGCGGCTGGCGGCG 549  
Db 104559 CVGGCGCGGNNHNNNSCGCGCGGCGGCGGCGGCGGCGGNNHNNNSCGCGGCGG 104500  
QY 550 AGCCCGCGGCGGCTGGCGAGCGGTGATGTACAGGCGAGCGGTGGGTCACTCG 609  
Db 104499 GCGCGCGGCGGNNHNNNSGCGCGGCGGCGGCGGCGGCGGNNHNNNSCGGCGG 104440  
QY 610 GAGGTAGGCGCGCGGCGGAGTTCAGCGAGTTCAGCGCGATTCGATTAGCAATG 669  
Db 104439 GCGCGCGCGGCGGCGGNNHNNNSCGCGCGGCGGCGGCGGCGGNNHNNNSCGGCGG 104380  
QY 670 AGCGCGCGGCTGGGTGGGTGTGTGTTAGGAGGAGGACCGCGGACCCCTCTCTC 729  
Db 104379 NNSCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGNNHNNNSCGCGGCGGCG 104323  
QY 730 CCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 789  
Db 104322 CVGGCGCGGNNHNNNSGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGNNHNNNSCG 104263  
QY 790 GGAGCTGGAGCGCGGAACTGGTACAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 849  
Db 104262 CCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104203  
QY 850 CCTCGAGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 887  
Db 104202 CCGANNHNNNSGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104165

## RESULT 9

AA53491  
ID AAX53491 standard; DNA; 114955 BP.  
XX AC AAX53491;  
XX DT 05-JUL-1999 (first entry)  
DE Human adenosine A1 receptor antisense oligonucleotide fragment.  
KW Antisense oligonucleotide; multiple target; antisense treatment;  
KW impaired respiration; inflammation; lung disease;  
KW pulmonary vasoconstriction; inflammation; allergic rhinitis;  
KW acute asthma; allergy; asthma; impeded respiration;  
KW respiratory distress syndrome; pulmonary pain; cystic fibrosis;  
KW pulmonary hypertension; pulmonary vasoconstriction; emphysema;  
KW chronic obstructive pulmonary disease; leukemia; lymphoma; carcinoma;  
KW colon cancer; breast cancer; lung cancer; pancreatic cancer;  
KW hepatocellular carcinoma; kidney cancer; melanoma; hepatic metastasis;  
KW prostate cancer; ss.  
OS Synthetic.

PN WO9913886-A1.  
XX PD 25-MAR-1999.  
XX PF 17-SEP-1998; 98WO-US19419.  
XX PR 09-JUN-1998; 98US-0093972.  
XX PR 17-SEP-1997; 97US-0059160.  
XX PA (UYEC-) UNIV EAST CAROLINA.  
XX PI Nyce JW;  
XX DR WPI; 1999-229400/19.  
XX PT New antisense oligonucleotides used in treatment of, e.g. pulmonary  
XX PT vasoconstriction  
XX FS Disclosure; Page 37; 120pp; English.  
XX CC The specification describes antisense oligonucleotides (AAX52869-X55271)  
CC directed against at least 2 mRNAs selected from target genes, coding and  
CC non-coding regions of RNAs corresponding to target genes, gene  
CC initiation codons, genomic flanking regions, intron-exon borders, the  
CC 5'-end, the 3'-end and the juxta-section between coding and non-coding  
CC regions and all segments of RNAs encoding proteins associated with one  
CC or more diseases, conditions or mixtures. The antisense oligonucleotides  
CC may be derived from sequences AAX55272-74. These multiple target  
CC oligonucleotides (specifically AAX55180-271) can be used for the  
CC antisense treatment of diseases and conditions. Typical diseases and  
CC conditions are those associated with impaired respiration and  
CC inflammation, including lung diseases, pulmonary vasoconstriction,  
CC asthma, allergic rhinitis, acute asthma, allergies, asthma, impeded  
CC respiration, respiratory distress syndrome, pain, cystic fibrosis,  
CC pulmonary hypertension, pulmonary vasoconstriction, emphysema, chronic  
CC obstructive pulmonary disease (COPD), and cancers such as leukemias,  
CC lymphomas, carcinomas e.g. colon cancer, breast cancer, lung cancer,  
CC pancreatic cancer, hepatocellular carcinoma, kidney cancer, melanoma,  
CC hepatic metastases, as well as all types of cancers which may metastasize  
CC or have metastasized to the lungs, including breast and prostate cancer.  
XX SQ Sequence 114955 BP; 6071 A; 29417 C; 36712 G; 21328 T; 21427 other;  
Query Match 5.4%; Score 54.4; DB 20; Length 114955;  
Best Local Similarity 30.0%; Pred. No. 0.015;  
Matches 213; Conservative 68; Mismatches 429; Indels 0; Gaps 0;  
QY 251 GGGCGGGCTTGTCTCGCGGAGCAGCAGCGTGGGAGGAGGCGCCCTCTGCGCTG 310  
Db 104079 GCGCGGCGGNNHNNNSGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104138  
QY 311 TGTGTTGCAACAGCAGCGCGCTGCGCGCTGCGGTTCCGCGCGCGGAGTCAACATG 370  
Db 104139 GCGTGGCGGCGGNNHNNNSGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104198  
QY 371 ATGTCACAGCAATGACACAGCGGTGTCTCTATTCACACAGCGTCCGAGCTGCACAA 430  
Db 104199 TCGGCGGCGGNNHNNNSGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 104258  
QY 431 TGTACACCGCGGTGCAACACATTTGGCCCGCGGAGCCCGGCGCTAGCCTCTGCGCG 490  
Db 104259 CGGCGCGGNNHNNNSGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 104318  
QY 491 CGCTCTCGCGGTCTCGGCGGAGTGGCCCGTTCCGCGCGGAGGCGGCTGGCGGCGCA 550  
Db 104319 CBGGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 104378  
QY 551 GCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 104378  
Db 104379 NNDNNSCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 104438  
QY 611 AGGTGAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 670

